

A clear case for closer European union

Britain's Labour government has given notice of its opposition to a new European 'superstate'. But the dangers should not obscure the advantages of mutual support in building up Europe's scientific infrastructure.

There is little doubt that the strength, if not the vitality, of US science is partly due to the way that 50 states of widely differing social and economic characteristics have been able to operate effectively as a single political unit. This has brought significant economies of scale to research. It has also given less advantaged states important help in their efforts to pull themselves up by the bootstraps, allowing them to draw on the support of federal agencies to promote infrastructures for scientific education and research.

This lesson should not be forgotten in Europe. There may be sound political reasons for remaining wary of pressures to create a superstate — Britain's new Prime Minister, Tony Blair, made clear last weekend in one of his first major policy announcements that he is unlikely to differ significantly from his predecessor, John Major, on this issue. But this does not undermine the fact that, as far as science is concerned, closer collaboration and mutual support promise substantial benefits, particularly (although not exclusively) to poorer member states of the European Union (EU).

Some of the evidence of these benefits is already clear. One is the way in which funding from the European Commission has helped to provide researchers from these countries with access to high-cost advanced research facilities which not even the larger countries could afford to construct on their own. Another has been the extensive use made of the EU's so-called 'structural funds' by some of the poorer states, such as Portugal (see page 115), to construct modern, well-equipped national facilities enabling them in principle to participate in international level science.

There are, inevitably, several dangers associated with the use of these funds to promote science. One, as some of the larger contributors to EU funds frequently point out, is that such funding may come to be seen as an alternative to domestic investment in research. This would

be a mistake, as it would divorce planning for research and development from other, equally important, dimensions of research policy (for example, the training of scientists and engineers).

A second danger, perhaps more serious for those working in such facilities, lies in the fact that the capital investment from Brussels includes no consideration for future running costs. This leaves the new institutions to compete for what are often highly limited funds with more established bodies, such as universities; some are already threatened with being stranded as expensive white elephants. And there is also the danger that the research programmes of such institutes, if planned outside regular scientific channels — perhaps influenced by strong local political factors — may lack adequate quality controls to justify the initial investment.

But all this calls for more European-level involvement, not less. One way that countries such as Portugal have been able to benefit from their membership of the EU is the opportunity that this has created to raise scientific standards by opening up research programmes to scrutiny by European peers. The more common agreement that can be reached across Europe on how investment in research can produce high quality, cost-effective results, the easier it will be to justify an increase in this investment, and the more productive it is likely to be.

Next week, Europe's research ministers meet to give their first opinion on the commission's proposals for the next multi-year Framework research programme. Inevitably, the ministers will face domestic pressures to interpret the concept of cost-effectiveness in a narrow, economic sense. But hopefully they will be sufficiently enlightened to accept that building what some have described as a 'scientific Europe' — in which all EU states are active participants and the richest feel a responsibility to extend a helping hand to the poorest — is both a desirable and a feasible objective. □

Red tape must not strangle good science

The reaction to a tritium leak on Long Island threatens US government laboratories with the last thing they need.

Federico Peña, the US energy secretary, has sacked the contractor at Brookhaven National Laboratory on Long Island, New York, accusing the management there of trading off the conduct of good research against the maintenance of high environmental standards (see page 114). Eliminating such a trade-off is certainly a laudable goal. But it is hard to believe that the dismissal of the Brookhaven contractor will do much to achieve it. More likely, it will reduce the prospects for the cost-effective conduct of top quality research by imposing yet more oversight, review and regulation on the Department of Energy's laboratories. That is the last thing they need, as Bob Galvin observed in his landmark report on laboratory reform.

The energy department's attempt to make up for past errors at Brookhaven involves the preparation of so many reviews and studies that it is hard to believe that anyone there will be left free to do anything else. It has also shut Brookhaven's research reactor, agreed to the early

departure of Nick Samios — as talented and communicative a director as any of its laboratories has ever had — and now sacked the contractor. All this because a monitoring project that Brookhaven planned in 1995 took place 18 months late, having failed to win enough priority points. As a result, a small — some would say insignificant — leak of tritium went undetected for that length of time.

It would have been too much to ask that Peña, on his visit to Long Island last week, should have sought to turn public attention to the diminutive nature of the public health risk posed by the Brookhaven leak. But, with brain research from the laboratory, published two weeks ago (see *Nature* 386, 827–833; 1997), adorning the latest cover of *Time* magazine, he did at least find time to praise the laboratory's scientific track record. Hopefully the performance of good science will remain the principal criterion by which the energy department judges management at its laboratories. □



Researchers warned: 'systems could crash on 1 January 2000'

[PARIS] Time is fast running out for research organizations that have not yet tested their information technology systems for the 'millennium bug' — a programming glitch in many microprocessors which will prevent them from recognizing the year 2000.

This, at least, is the growing consensus among computer experts, who predict mayhem in many sectors of society, including parts of the research community, unless computers and intelligent devices are repaired before then. Estimates of the worldwide costs of doing so are as high as US\$600 billion.

Many research organizations are beginning to take the problem seriously. But others continue to give what some claim to be insufficient attention, pointing out that many scientists seem allergic to the hype generated by the thousands of consultants and computer companies feeding off the problem.

The problem originates from the practice of programmers representing years as two digits (dd/mm/yy) instead of four (dd/mm/yyyy) in many operating systems, software packages and electronic devices. As a result, many microprocessors will reset their clocks to 1900 in the year 2000, a date now known as 'y2k'. Any operation that adds, subtracts or compares dates may generate errors, or, at worst, cause systems to crash.

But the glitch is not restricted to computers. It also affects embedded microchips, which are found everywhere from toasters and lifts to laboratory equipment and air-

craft. These chips often carry date functions.

The major concern for research organizations is in administration of such activities as payroll, procurement and grant tracking. "Our [business] software is supposed to be y2k-proof, but nothing is being left to chance and a full simulation is being planned," says Michael Metcalf, information technology specialist at the European Laboratory for Particle Physics (CERN) in Switzerland.

The situation there is comparable to the problem facing organizations such as banks and insurance companies which are highly dependent on data processing, and so most exposed to the y2k problem. Much of business still runs codes written in the 1960s and 70s using COBOL, a language that used two-digit dates, while even more recent software often contains chunks of older programs. Programs often run to millions of lines of code, all of which may need to be checked.

"Technically the y2k problem is easy to fix; it's the scale of the problem that makes it difficult," says Martin Jourdain, a researcher at the French computer research agency, INRIA, who is on secondment to Metaware, a company dedicated to the y2k problem. "Dates can hide in places you never suspect."

IBM estimates that remedying the problem worldwide will require 1,910,000 programmers. COBOL programmers are now commanding salaries double those 12 months ago.

Scientific computing seems to be much less vulnerable to the y2k problem than

research administrations. Operating systems and applications are written in scientific and engineering languages where dates are usually stored as 32- or 64-bit timestamps. The 32-bit Unix datestamp used by Cray Research has measured time in microseconds since the birth of Unix in 1970, for example, while the most recent Crays use a 64-bit timestamp with a lifespan of 2^{34} years. But this does not prevent programmers from employing problematic date formats in programs.

Kent Koeninger, head of Cray's Year 2000 programme, says supercomputer codes are revalidated every time they are transferred to a faster machine and that correcting date problems has been part of this exercise. Cray has found minor y2k problems but nothing critical, he says. "We are taking the problem seriously, but we have much less work to do than in the data processing shops."

CERN will be shut down for Christmas 1999 until 3 January 2000, so "no circulating beam will risk appearing to be a century old", says Metcalf. "If anything untoward happens when tests start later in the year, this is anticipated to be no worse than the biannual change to and from summer time."

The y2k problem facing the US space agency NASA is "not trivial, but not all that pervasive for us," says Jim Radosevich, coordinator of NASA's \$30-million Year 2000 programme. He says most of the problems concern administration, and desktop machines using commercial software.

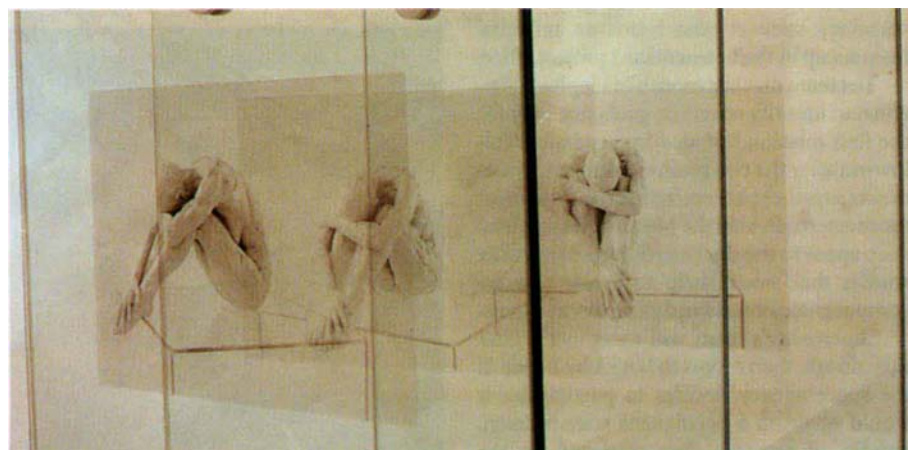
Space science benefits from the fact that missions are timed as the number of hours from launch and not with respect to a specific date, he says. "On the scientific side, there is potential for slight problems but we haven't found anything serious in mission-critical systems."

US federal agencies estimate that they will spend \$2.3 billion altogether between 1996 and 2000 on y2k. A report from the US Department of Energy foresees "a potential national emergency that can seriously degrade or threaten national security. If the problem is not addressed, the impact to the federal government could be insurmountable."

The report's list of mission-essential systems that must be checked includes "safe operation of nuclear power production facilities, control and accountability of nuclear materials, control of nuclear facilities; health protection from hazardous materials; and medical treatment of patients."

Many scientists remain sceptical, however, of the claims that many systems will melt-down in 2000 if left uncorrected. "This is all hype and there is a bunch of people that want

Art and science shape up for awards



[LONDON] Six collaborations between artists and scientists in Britain have been awarded a total of £90,000 (US\$146,000) under SCI-ART, a science and art initiative launched by the Wellcome Trust. Arthur Crisp, of Atkinson Morley's Hospital in London, and

artist Shelley Wilson have received £12,500 for a project entitled *Portraits of Anorexia* that will produce artwork conveying this disorder to the public and showing how body shapes are categorized. An example of Wilson's earlier work is shown above.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Will the 'millennium bug' cause microchip mayhem — or is the scare just hype?

to make a lot of money out of saying there is a big problem", says John Crowcroft, professor of network systems at University College London. Crowcroft adds that tests of the core Internet protocols and applications for y2k problems "found a couple of things where the user might be misled, but otherwise the system functioned fine".

The large scientific databases are also taking a relatively relaxed stance. "I'm not profoundly alarmed," says Graham Cameron, head of services at the European Bioinformatics Institute in Cambridge, England. The institute's systems do not use two-digit dates.

Many UK universities are ignoring the problem, claims one member of the UK government's Year 2000 Task Force. The French atomic energy commission (CEA) has no plans to establish a formal Year 2000 programme, according to a spokeswoman, Corrine Borrell, who says the agency is content with raising awareness among staff.

But scientists who play down the problem are being "extraordinarily irresponsible", asserts Robin Guenier, a consultant who heads the UK task force, and is a former chief executive of the government's Central Computing and Telecommunications Agency.

He says: "This is not being exaggerated. If anything it is being underestimated, and there is little time to fix it." Anthony Parish, director general of the UK Federation of Electronic Industries, agrees: "We think it is a major disaster; it will cost a large amount of money."

One reason for such concerns is growing awareness of the problem of embedded microchips. "These are beginning to be acknowledged as perhaps a bigger problem [than software]," says Simon Reeves, co-author of a recent book called *The Millennium Bomb*. Pharmaceutical companies are reported to be highly concerned about the chips used throughout process plants.

But claims that embedded chips will cause widespread problems, such as shut-downs of power plants and aircraft, are contested by several scientists. Darrell Ince, a specialist in critical systems software at the UK Open University, argues that such problems would be localized and rarely cause crashes. "I don't think you will get problems in safety-critical systems such as planes flying upside down."

Declan Butler

Space Shuttle launch holds key to asteroid mission

[WASHINGTON] A controversial US military spacecraft mission to fire a probe into an asteroid's surface may finally get off the ground in 1999 if Congress commits additional funding for next year, and if the National Aeronautics and Space Administration (NASA) offers a free ride on the Space Shuttle.

Managers of the Clementine 2 project hope the Air Force will approve their mission plan as early as next week, which would allow work to resume at the Naval Research Laboratory in Washington DC and at the Air Force Phillips Laboratory in New Mexico. The project has been in abeyance for much of the past 18 months due to budget politics at the Pentagon (see *Nature* 381, 8; 1996).

One result of that controversy is that Clementine 2 is no longer seen as part of a broad "planetary defence" mission to protect the Earth from rogue asteroids. The rationale was always considered suspect in some sectors of the Pentagon, and so advocates of the project now emphasize its advanced technology.

Like the Clementine 1 mission that mapped the Moon in 1994, the project is primarily a test of miniature spacecraft components, cameras and sensors developed for space-based missile defence. The plan calls for the spacecraft to launch in November 1999 and fly past two asteroids — 1986JK and Toutatis — in 2000. A probe would be sent crashing into each one, stirring up material that could be studied by spectrometers and cameras on the main craft.

NASA's involvement in the project is crucial, to "show that there really are valid and exciting scientific objectives", says Eugene Shoemaker, a retired US Geological Survey planetary scientist who heads an informal team set up by the Clementine 2 project office.

The team met last month in Flagstaff, Arizona, to identify scientific goals not only for the first mission but also for a possible follow-on after the two main encounters. Shoemaker says the spacecraft could go on to meet more asteroids, visit the Moon or be sent into deep space to conduct astrophysical parallax studies that would help to locate massive compact halo objects and gamma-ray bursts.

Shoemaker's team will meet again later this month, then report to NASA by 1 June. If the space agency decides to participate, it would establish a permanent science team, handle all research data returned by the spacecraft, and perhaps contribute an instrument in addition to a suite of cameras and spectrometers already under development at the Lawrence Livermore National Laboratory in California.

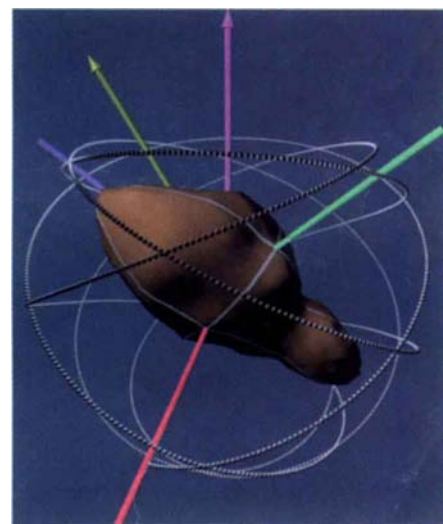
A second key question is whether NASA

will agree to launch Clementine 2 on the Space Shuttle at little or no cost. The shuttle programme chief, Stephen Oswald, said recently in congressional testimony that NASA is interested in providing a launch, but no details have been worked out. Part of the problem, say Clementine supporters, is that until now NASA has been unsure that the Air Force is fully behind the project. Approval of the Clementine 2 mission plan, and the resumption of work at the Naval Research Laboratory and Phillips, should help to ease those doubts.

The final ingredient necessary for Clementine 2 to fly is more money from Congress. The project received only \$45 million last year — half the amount it requested — as the result of a political deal between Republicans and Democrats to trim the defence budget. Without an additional \$50 million in 1998, the asteroid fly-bys are out of the question, and the project would be scaled back just to test the probes in Earth orbit. Clementine supporters are optimistic that the Senate Armed Services Committee, which authorizes defence spending, will add the money, and that appropriations committees will follow suit.

Clementine is not the only science project being negotiated between NASA and the Air Force. The two agencies are close to finalizing an agreement whereby the Jet Propulsion Laboratory would provide upgraded cameras to three space-tracking telescopes used by the military. The Air Force would gain from the added camera performance, and the space agency would be able to speed up and improve its search for near-Earth asteroids.

Tony Reichhardt



In a spin: Clementine 2 will fire a probe into the Toutatis asteroid (above) if it can hitch a ride on the Space Shuttle.

Germany gets warning on access to sequence data

[MUNICH] Germany's efforts to improve its standing in the international genetics community are threatened by a controversy about rules giving German industry three months' privileged access to the results of research funded by the government through its human genome research programme.

The research programme was launched by the research ministry in 1995. It has recently been agreed that all researchers receiving support through the programme will in future have to make their data available to a new German industrial funding association, Verein zur Förderung der Humangenomforschung.

The idea is to give German drug companies involved in the association three months' privileged access to the data. But the proposed rule has led to a threat from the directors of several large genome sequencing centres in the United States, the United Kingdom and France that they will block access to their sequence data for German scientists and industry.

In response to the threat, the research ministry says it will suspend the agreement with the association until the issue is resolved. A meeting will be held in Bonn at the end of the month with representatives of the association and the international academic community to try to resolve the conflict. Until then, German researchers will remain free to publish and distribute their data as they wish. But they have been forbidden to disclose details of the conflict.

The association was set up by German industry in response to criticism that industry had declined to contribute to the human genome programme, which was intended to be financed equally by the government and industry to a total of DM400 million (US\$232 million) over four years (see *Nature* 378, 6; 1995).

In contrast, several companies have forged bilateral links with genomics companies in the United States. Hoechst last month formed a US\$85-million joint venture with Ariad Pharmaceuticals in Cambridge, Massachusetts, to develop small molecules to regulate gene function in target diseases.

The new association includes the drug companies Asta Medica, Bayer, BASF, Boehringer Ingelheim, Boehringer Mannheim, Hoechst, Merck and Schering. Each is understood to have agreed to contribute DM150,000 a year.

The money will be used primarily to support a Patent and Licensing Agency. This will have three months to review findings resulting from ministry-funded research and, where considered appropriate, to file patents

on behalf of researchers. The agency will pay the filing costs, and any member of the association can then bid for the patent rights.

But the plan drew an angry response from the scientific community when it was unveiled by Ursula Hurtenbach, who organizes the human genome research programme for the research ministry, at an international strategy meeting on human genome sequencing in Bermuda in February.

At the first such meeting, held in Bermuda in 1996, most of the publicly funded human genome centres represented agreed to the principle that all genetic information should be made freely available to scientists and industry (see *Nature* 371, 365 & 551; 1994).

Participants at both meetings included John Sulston, director of the UK Sanger Centre, Jean Weissenbach from Génethon in Paris, Michael Morgan of the Wellcome Trust in London, and Francis Collins, director of the National Center for Human Genome Research at the National Institutes of Health in Bethesda, Maryland.

They told Germany at this year's meeting that its research institutes would be excluded from the previous consensus agreement if plans for the new association were not changed. All four participants have been invited to take part in the Bonn negotiations, which will be chaired by Josef Straus, director of the Max Planck Institute for Foreign and International Patent Copyright and Competition Law in Munich.

Many German scientists are worried at the prospects of being excluded from the agree-



ment. "If the research ministry and the association continue to insist on privileged access, scientists and industrial participants are likely to lose access to 90 per cent of the genomic sequence information generated by international centres", says André Rosenthal, head of the department of genome analysis at the Institute for Molecular Biotechnology in Jena. Rosenthal expects to receive a grant from the ministry of around DM20 million for a three-year project to sequence 40 megabases on human chromosomes 21 and X.

But German companies are maintaining their hard-line stance. "We will never support research whose findings are distributed indiscriminately," says Rolf Krebs, a member of Boehringer Ingelheim's executive board.

The companies also argue that the US patent system gives a considerable competitive edge to US industry because it allows a 'grace period' under which patents can be applied for up to one year after the publication of research results. To compete effectively, the companies argue, European industry needs privileged access to the results of European research.

Robert Unterhuber

Rüttgers and Chirac seek cloning ban

[PARIS] Germany's research minister, Jürgen Rüttgers, last week backed a call by the French president, Jacques Chirac, for a worldwide ban on human cloning. Chirac, who made the statement on the same day he received a commissioned report on the subject from the French national bioethics committee, declared his "vehement, categorical and definitive ethical condemnation of all reproductive cloning of human beings".

Rüttgers also presented a statement from seven leading German scientific experts on the biological,

ethical and legal aspects of cloning. He said he agreed with their conclusion that the cloning of humans is ethically unjustifiable, and called for the rapid introduction of an international ban. In Germany, human cloning has been illegal since 1990, the maximum punishment being five years in prison. France's 1994 bioethics laws also ban human cloning.

"Cloned humans in any part of the world would mean an attack on the dignity and integrity of every single man on Earth", said Rüttgers, adding that the

Council of Europe's convention on bioethics should be modified to ban cloning explicitly. Chirac said he would raise the issue at the next summit of the G7 countries in June. He suggested that the International Bioethics Committee of the United Nations Educational, Scientific and Cultural Organization, which is chaired by Noëlle Lenoir of France's constitutional council, was the "natural forum" for introducing an international ban.

Declan Butler &
Quirin Schiermeier

ESA economies may allow Mars mission

[MUNICH] The European Space Agency (ESA) is drawing up plans for a quick, low-cost mission to Mars as a way of partly restoring the frequency of research missions, which are under heavy pressure from falling budgets.

ESA's Space Science Advisory Committee discussed the plan for the Mars Express mission last week. But Lodewijk Woltjer, the head of the committee, warns that the mission will be possible only if the agency's member states do not impose further financial restrictions on the space science programme.

The committee of eight European space scientists had been asked to redesign ESA's long-term Horizons 2000 plan in the light of the programme's reduced financial projections. Horizons 2000 is a mixture of large cornerstone missions and medium-sized missions designed to balance the needs of Europe's space science community.

In January, the committee proposed stretching out the programme. But this suggestion came under criticism from some member states because delaying launches would disrupt the flow of data to European scientists in a way that they found unacceptable (see *Nature* 386, 421; 1997).

Following a series of studies into ways of

saving costs, the committee is now proposing that two planned missions should be launched together in 2005 or 2006 — the cornerstone Far Infrared Telescope, and the medium-sized Planck Surveyor, which will study microwave emission from the sky. By sharing the launch, the spacecraft and the high-technology cooling system, ESA expects to save ECU300 million (US\$340 million), without significantly delaying either mission.

The committee is proposing that half of the sum saved should help to pay for the relaunch of Cluster, the solar science mission that was destroyed when its Ariane 5 launcher exploded last year. The other half, it says, should be reserved for the Mars mission.

The committee says the costs of Mars Express can be kept down to around ECU150 million by using blueprints of instruments designed for other missions but never used. These instruments, including some of those lost on the Russian Mars 96 mission when its Proton launcher failed last November (see *Nature* 384, 199; 1996), would allow some fundamental scientific experiments to be carried out, such as unique spectral analysis of the surface of Mars.

Woltjer stresses that there are two reasons

why the proposed 2003 launch date must be kept to if the mission is to stay within its budget. First, the positioning of planets during that year will leave a clear path to Mars, minimizing the fuel required. Second, the scientific relevance of the instruments could be reduced after this date by the continuing Mars exploration programme.

The committee is also proposing a significant future participation in the Next Generation Space Telescope, the follow-up to the Hubble Space Telescope.

The cost-saving analyses by ESA have been unable to identify how to bring forward the next unallocated medium-sized mission, which remains delayed until 2007. Nor could the agency fix a date for the next cornerstone mission, originally due to be launched in 2009. This was intended to be an interplanetary mission to Mercury but the committee is now to reconsider its aim.

A series of cheap technology testing missions, SMART (Small Missions for Advanced Research in Technology), is expected to get off the ground. Officials propose that the first, to demonstrate the feasibility of solar electric propulsion as a way of reducing fuel requirements for interplanetary missions, should be launched to the Moon in 2002. **Alison Abbott**

US science adviser's exit 'may herald new White House strategy'

[WASHINGTON] Jack Gibbons, science adviser to President Bill Clinton, has countered growing speculation about his future in the post in a statement saying that he is staying put — but "may desire to leave near the end of the year".

Gibbons issued the statement in response to reports that John Deutch of the Massachusetts Institute of Technology was in line to replace Gibbons this summer. Deutch has held various senior posts in a series of recent administrations, mostly recently as director of the Central Intelligence Agency (CIA).

The vague statement has left science lobbyists in Washington concerned that uncertainty about Gibbons' future could undermine the influence of the Office of Science and Technology Policy (OSTP), which he runs in the White House.

But well-placed sources say that the statement points to Gibbons presiding over a transition period before making way, probably at the beginning of next year, for a new adviser who, it is said, will increase the profile of science and technology issues at the White House.

Administration officials and Deutch have declined to confirm or deny that he will be that person — fuelling speculation that he is the front-runner. The statement said: "Dr



Gibbons: doubts over his effectiveness.

John Gibbons continues to serve as the director of OSTP, and continues to advise the president. Any contemplation about his successor is premature speculation. Dr Gibbons has indicated that he may desire to leave this post

near the end of the year. When that time comes, a number of leaders in the US science and technology community will be considered for this critical position."

Reports of a Deutch appointment surfaced as science advocates expressed growing concern about the lack of a clear statement about Gibbons' intentions, as well as doubts about his effectiveness in Washington's corridors of power. One senior physicist said privately last week that Gibbons was "not a player" in Washington. And an official in a major scientific society said OSTP was "adrift", having "a director whom everyone says is going to leave".

After the statement, science lobbyists said that it might make things even worse. If a successor is not in place until 1998, they said, that person will have little time to accomplish anything before the second Clinton term,

which ends in 2000, enters the "lame duck" phase which besets most presidents in their final 18 months in office.

Gibbons refused to expand on his statement, or to comment on reports that OSTP's influence in the White House was being usurped by John Podesta, Clinton's deputy chief of staff and a former lobbyist for high-technology corporations. According to the newsletter *Washington Fax*, Podesta wants the administration to restore the strong emphasis on technology support that Clinton promised early in his first term.

The Deutch rumour puzzled officials in the Congress. They say that Deutch left the CIA job on bad terms with Clinton, after he had failed to land the job he wanted as secretary of defence. They point out that the science adviser's job, as traditionally defined, is junior to both of these positions.

But the appointment of a high-profile adviser such as Deutch could fit in with a long-term White House science policy strategy. This was outlined by one official, who said that Gibbons could be a "transitional figure" from that of a traditional science adviser to one under which science would become a "major policy area". Such a move would place concern for research on a par with economics or national security. **Colin MacLachlan**

Europe sets out plans for solving neutron drought...

[MUNICH] Five European research institutes have agreed to carry out a three-year research and development assessment of what a technical study published last week says would be the world's most powerful neutron source — the European Spallation Source (ESS).

Planned to run at 5 MW, the ESS would be 30 times more intense than the ISIS facility at Rutherford–Appleton Laboratory in the United Kingdom, the most powerful operational spallation source. But it would not be cheap — construction costs are estimated at ECU935 million (US\$1,062 million), and annual running costs ECU85 million.

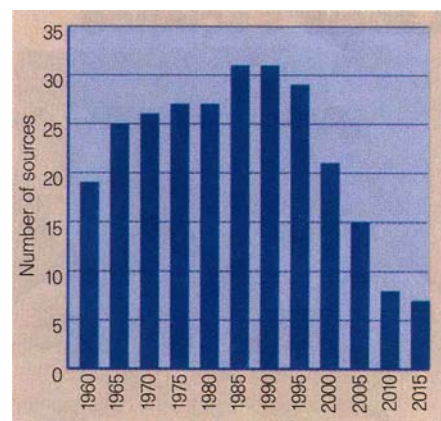
Plans for such an ambitious project emerged following warnings of a European and global 'neutron drought' within the next decade, as demand for neutrons grows and many existing neutron sources come to the end of their working lives (see chart).

This threat was identified by a European Commission study in 1990, and by the Organization for Economic Cooperation and Development's Megascience Forum in 1993. The studies recommended the immediate development of next-generation neutron sources (see *Nature* 379, 284; 1996).

Scientific workshops set up by the Megascience Forum to investigate the problem concluded that the new-generation neutron source should be based on spallation technology, where neutron beams are produced by bombarding a heavy-metal target with high-energy particles from a powerful accelerator. This is generally recognized as having greater potential than the traditional nuclear fission method, where the high level of heat produced limits the maximum flux of neutrons.

Last week's technical study was published by the ESS Council, an international group of experts supported by the European Commission. It estimates that the ESS would be able to produce six to eight times as many neutrons per unit energy produced as could a research reactor. Neutron beams from a spallation source can be pulsed, which gives them greater scientific flexibility.

The new ECU20-million research and development assessment will be conducted by scientists from the Commissariat à l'Énergie Atomique in France, the Council of the Central Laboratory of the Research Councils in the United Kingdom, the Research Centre Jülich in Germany, the Paul Scherrer Institute in Switzerland, and the Risø National



The number of neutron sources in OECD countries is due to fall radically.

Laboratory in Denmark. It will complete engineering data but will also look at cost-saving options such as the possibility of developing superconducting technology to reduce running costs.

Participating institutes will help to pay for the study, but money is also being sought from other sources, including the commission's Training and Mobility of Researchers programme. Jørgen Kjems, chairman of the ESS Council and head of the Risø laboratory, says a mobility grant would have the advantage of training young scientists to take part in what could be a long future for the project.

Despite claims that the ESS, if built, would become the world's largest neutron source, neither Japan nor the United States is keen to concede the lead in neutron source development. Both are planning similar projects, and are working to slightly shorter timetables.

If the ESS finds support for construction, which might include funding from Brussels to allow participation of smaller European Union states, it is likely to start serving the European community of some 4,000 neutron-scattering scientists no earlier than 2010.

The United States is planning a 1-MW National Spallation Neutron Source (see panel) which could be upgraded to 5 MW. Japan is planning two spallation sources: the 1-MW KEK Japanese Hadron Project in Tsukuba, scheduled for 2003, and the 1- to 5-MW JAERI Omega, scheduled for 2008.

Despite the rivalry, much collaboration between scientists in these projects is likely. "This is a very positive outcome of the Megascience Forum workshops," says Kjems. "Europeans, Americans and Japanese scientists will share their experiences in, for example, the development of suitable targets, and this will keep down costs."

Last week saw the signing of a memorandum of understanding between the five research centres. Details such as how the ESS could be funded and where it would be sited are not yet being addressed. Scientists hope research ministers will look favourably on financing when discussing the forum's report on neutron sources next year.

Alison Abbott

...as US looks again at Oak Ridge site

[WASHINGTON] A review team of 60 experts will descend on Oak Ridge National Laboratory, Tennessee, next month to assess its proposal for a new spallation neutron source which could propel the United States back into the lead in neutron science.

Officials from the European Spallation Source will be invited to join the reviewers, whose judgement will help to decide the construction schedule for the proposed machine. Known as the National Spallation Neutron Source, it would be six times more powerful than ISIS at the Rutherford–Appleton Laboratory.

But many obstacles remain. The Department of Energy is short of money for new projects, and there are doubts about whether Oak Ridge is an appropriate site. Construction cannot start before 1999, and is unlikely to be finished before 2004.

The machine is expected to cost \$1 billion. The department says it should be built at Oak Ridge partly because the laboratory lost out on an earlier, \$3-billion proposal for a neutron-producing reactor that was cancelled in 1995. Vice-President Al Gore pledged to build the spallation neutron source at Oak Ridge during last year's election campaign (see *Nature* 383, 207; 1996).

But the Science Committee of the House of Representatives has recently proposed legislation requiring the energy department to ask the National Academy of Sciences to compare this choice with rival laboratories before proceeding with the project.

Users of other US neutron sources are increasingly worried about slow progress in upgrading them. The American Physical Society has issued a statement

expressing its concern.

A small upgrade to the High Flux Isotope Reactor at Oak Ridge will provide 'cold' neutrons later this year. And a \$36-million upgrade to the Los Alamos Neutron Scattering Experiment is expected to provide a 160-kW short-pulse spallation source at the New Mexico laboratory by 2001. But a larger upgrade at Los Alamos, to provide a long-pulse source, has not yet been approved.

Worst of all for US neutron scientists, the High Flux Beam Reactor at Brookhaven National Laboratory in New York — which also needs refurbishment — will remain closed at least until December. Earlier this year, radioisotopes were found to have leaked from the pool that stored its used fuel (see *Nature* 386, 3; 1997).

Colin Macilwain

Labour's science appointment lifts hopes for reform

[LONDON] Britain's new Labour government has appointed John Battle, opposition spokesman on science and technology from 1994 to 1995, as the minister responsible for science within the Department of Trade and Industry. He will report to Margaret Beckett, who speaks for science in Cabinet as President of the Board of Trade.

Battle, a geography graduate, is associated with the 'soft left' in the party — placing him between the traditional 'old left' and the reforming 'new Labour' of Prime Minister Tony Blair. He demonstrated an enthusiastic interest in his portfolio during his period as opposition spokesman, and his appointment is likely to be welcomed by those in the scientific community who had feared that science could end up in more conservative hands.

It could, for example, signal some significant changes in science policy. Two years ago, Battle announced after almost 50 visits and discussions with key scientists and research institutions that two "primary themes" had "clearly emerged" — the need to break with short-termism, and the need for a "nationally co-ordinated science strategy". The lack of such a strategy, he argued, had resulted in a reduction of government support for science.

The latter conclusion is similar to that reached by the House of Commons Select Committee on Science and Technology in its recent inquiry into the research councils, suggesting that the topic is likely to be high on the new government's agenda.

Battle also expressed a desire to consult widely with the scientific community, announcing plans for a series of regional public meetings described as a "massive science policy consultation exercise" known as the Science 2000 Project. But this plan was dropped soon after his successor, Adam Ingram, who has been appointed to a ministerial post in the Northern Ireland office, took over the science portfolio.

Since 1995, Battle has been Labour's energy spokesman, and he will keep responsibility for energy — as well as for information technology — in his government post. Although some are concerned that political responsibility for science has been diluted as a result, Battle will be a full minister in the department. In contrast, his predecessor, Ian Taylor, while concentrating solely on science and technology, occupied a rung down the political ladder as a parliamentary under-secretary of state.

Within the Department of Education and Employment, responsibility for universities will be handled by Baroness Tessa Blackstone, the master of Birkbeck College in London, who has been appointed minister for higher education.

David Dickson

Brookhaven contractor is sacked over tritium leak

[WASHINGTON] The US Department of Energy announced last week that it was sacking Associated Universities Incorporated (AUI), the contractor that has run Brookhaven National Laboratory on Long Island, New York, for 50 years, after an internal report criticized management and environmental procedures at the laboratory.

Department officials also said that the High Flux Beam Reactor (HFBR) — the research reactor whose fuel storage tank has leaked small quantities of tritium, plunging the laboratory into crisis — will not reopen until the local community is happy that it can operate safely (see *Nature* 386, 3; 1997).

Tara O'Toole, assistant secretary for environment, safety and health at the energy department, says that HFBR's users shared the blame for the crisis at Brookhaven. "It is wrong-headed to say that the users are not at fault," she said, when asked whether scientists who use HFBR as a neutron source would get money to move to other facilities.

"The users have to take responsibility — that is the message we have to get through to scientists," says O'Toole. Environmental health and safety "is not a separate activity" for non-scientists at the lab to deal with.

Federico Peña, the energy secretary, went to Brookhaven to release the report and fire the contractor last Thursday, 1 May, only four days after the new president of AUI, Lyle Schwartz, had been appointed as interim director of the laboratory. Schwartz, who left a senior position at the National Institute of Standards and Technology to join AUI in March, says he will remain interim director at Brookhaven until a new contractor is appointed in six months.

Peña said that AUI was being sacked for "unresponsiveness" to his department's "needs and expectations for community relations and environmental safety and health stewardship". But Paul Martin, chair of AUI and dean of engineering at Harvard University, said that the decision was "taken precipitously".

Appointment of a new contractor at Brookhaven will probably mean a clear-out of top management, as well as a new employer, and perhaps new terms and conditions, for the 3,200 staff.

AUI was set up in 1947 to run Brookhaven. Its other main activity is the management of the National Radio Astronomy Observatory for the National Science Foundation. The observatory has recently experienced contractual problems in building a \$75 million telescope at Green Bank, West Virginia (see *Nature* 384, 505; 1996).

The AUI board of trustees will meet in Washington this week to decide whether to



Under a cloud: users of the Brookhaven National Laboratory (above) 'must share blame'.

bid for the Brookhaven contract — as it is theoretically entitled to do — and, Schwartz said, to assess its overall future.

John Wagoner, the energy department official currently responsible for cleaning up the dirtiest nuclear waste site in the United States at Hanford, Washington, has been seconded temporarily to Brookhaven, from where he will report direct to secretary Peña. "John has proven under fire that he's a leader," says O'Toole.

Wagoner concedes that the environmental problems at Hanford and Brookhaven are "different by orders of magnitude". Brookhaven has detected a total cumulative leak of 5 curies of tritium, while Hanford discharges 6,000 curies of radioisotopes into the Columbia river each year and has around 200 million curies of waste stored in containers of varying integrity.

Peña announced that a team of inspectors from the Environmental Protection Agency is due to arrive at Brookhaven this week to perform a "full facility inspection". And Martha Krebs, assistant secretary for research at the energy department, has been given 30 days to draw up a fresh "action plan" for environmental and safety management and community responsiveness at the laboratory.

The upsets at Brookhaven are now set to spill over into the rest of the department's laboratory complex. According to Krebs, "this is an issue not just for Brookhaven but for all of our multiprogramme laboratories".

O'Toole said last week that the management of Brookhaven had "lost the public trust" and accused the lab of "problematic attitudes" to environment, health and safety issues. An example of these attitudes, she said, was that electricians carried out maintenance work on an accelerator facility at Brookhaven while wires were still live, to avoid interrupting scientific experiments.

Colin MacIlwain

Science in Portugal joins the mainstream

David Dickson

The purposeful use of funds from the European Union has provided a much-needed fillip to the research enterprise in Portugal. But with progress comes a range of problems familiar elsewhere in Europe.

[LISBON, PORTUGAL] Over the next few weeks, the Portuguese government will put the finishing touches to a series of reforms designed to help equip the country with the science infrastructure needed to help its economy catch up with those of its European partners.

In financial terms, Portugal has a long way to go. Its total spending on research and development, standing in 1992 at 0.63 per cent of the gross domestic product, is less than that of any other member of the European Union. For example, its nearest neighbour, Spain, spent 0.91 per cent.

But the socialist government that came to power in December 1995 is determined to change this situation. The latest reforms — including the creation of a new national foundation to fund basic research — are intended to streamline science in the universities and research institutions.

In financial terms, research, together with education, is already a priority; in an otherwise austere budget, funding for research, which was kept constant in the early 1990s, has been increased by 16 per cent this year.

European dimension

Central to the reforms has been Portugal's integration into the rest of Europe, a process that started with the 'revolution' of 1974, which freed the country from the dictatorship of Antonio de Oliveira Salazar, and which has accelerated dramatically since Portugal became a full member of the European Union in 1989.

It is therefore appropriate that the reforms are being led by José Mariano Gago, the minister for science and technology. Before becoming chairman of Portugal's national research funding committee in 1986, he was a research physicist at the European Laboratory for Particle Physics (CERN) in Geneva.

Part of the Europeanization of Portuguese science is clearly visible in the substantial use the country has made of so-called 'structural funds' — provided by Brussels to boost the economic development of the poorer parts of the community — to build a range of large, modern research facilities. Access to these funds was one of the concessions obtained by Portugal in joining the European Union in exchange for opening up its economy. While some countries have used them for more tra-

ditional projects, such as building roads and hospitals, Portugal, which had hardly any domestic science-based industry, has used them to boost its scientific potential.

A second element in this Europeanization process is Gago's determination to ensure a high representation of foreign scientists on 20 evaluation panels. These have been set up to assess the performance and plans of more than 300 publicly funded research units based in the country's universities and other institutions.

The results of the evaluation, which were sent out to institutions last month, are being used by the government as the basis of future support (15 of the 300, for example, have already been told their funding will end).

Gago says that the participation of scientists from other European countries has been central to Portugal's success in focusing funds on productive laboratories. "It was essential to have the involvement of foreigners in the evaluation panels," he says. "This practice should be built into European science policy and into bilateral agreements."

This process of rationalization and selectivity has been reflected in changes in other research organizations. Perhaps the most prominent is the Gulbenkian Institute of Science, founded on the legacy of the legendary oil entrepreneur Calouste Gulbenkian, whose research laboratories at Oeiras outside Lisbon have long been one of the country's few centres of international scientific excellence.

The foundation has recently introduced sweeping changes, in preparation for the arrival of its new director, António Coutinho, currently at the Institut Pasteur in Paris. The changes have not been without their critics, particularly workers nearing the end of their research careers who are being asked to make way for younger researchers.

But Horácio Menano, the foundation's director, argues that change is essential if the institute is to maintain scientific vitality. Furthermore, the new plans envisage the institute



Gago: optimistic about the future.

becoming considerably more international in its activities, offering a temporary location for scientists around the world.

The broader changes in Portuguese science have not been without criticism, even from those who support their direction. Some, for example, claim that the research institution evaluations could have been applied even more rigorously, while the quality of the review process was itself mixed. Others have difficulty with the way in which structural funds from Brussels have been used to create many independent research entities — often set up to avoid the bureaucratic complexities of channelling funds through more conventional bodies.

Even those who have benefited from such funds claim they now face major difficulties because, while the money received can cover construction and equipment costs, there is no additional funding for operating costs; this must come from the national government. "At least when you have built a road, you can then charge tolls on it," says Antonio Xavier, director of the Institute of Technology for Chemistry and Biology, which was built near the Gulbenkian Foundation.

Climate of change

Despite such difficulties, combined support from Lisbon and Brussels has, since the early 1990s, created a research-friendly environment in Portugal that is already beginning to bear fruit. This is reflected in the growing number of Portuguese scientists (such as Xavier) who have returned from abroad to senior posts.

"The whole climate for science in Portugal has changed for the better," says Alexandre Quintilhas, who returned from 20 years at the Lawrence Berkeley Laboratory in California to head the Institute for Molecular and Cell Biology, also built with structural funds at the University of Porto in northern Portugal.

According to Quintilhas, the main problem facing Portuguese science — apart from inadequate funding from the domestic budget — is that of providing adequate career opportunities for postgraduate students.

Gago says that training and employment of postdocs is "one of our top priorities now". One scheme launched recently by the government will pay on average half the salary of a PhD employed in industry for the first three years.

Many postdocs continue to face uncertainty futures, particularly given the lack of new openings in universities. Gago says that, at present, the creation of new institutes means the country has no problem finding work for its PhDs, although he admits that "the number of people in non-permanent jobs after their PhDs is increasing".

Overall, however, Gago is optimistic about the future. "Science in Portugal is becoming much more normal in international terms," he says. □

Clinton offers to say sorry for Tuskegee Syphilis Study

[WASHINGTON] President Bill Clinton is to apologize for the notorious 'Tuskegee Syphilis Study' in which the US government withheld treatment for the lethal disease from 399 black men in order to follow its progression and the mechanism of death. The study was conducted by the Public Health Service between 1932 and 1972.

Clinton will apologize to the families of the men and the eight surviving victims at a White House ceremony on 16 May. Michael McCurry, a Clinton spokesman, said the study was "absolutely reprehensible".

It is not clear whether the survivors will attend. Their lawyer said they would prefer that the president apologized in Alabama county where the study was conducted.

Moniz tipped for energy department post

[WASHINGTON] Ernie Moniz, chair of the physics department at the Massachusetts Institute of Technology, may be about to join the Department of Energy as under-secretary, the third most senior position in the department. Federico Pena, the new energy secretary, says that he has selected

someone to fill the post, vacated earlier this year by Tom Grumbly. Several sources say that the person is Moniz.

Pena had wanted to appoint a top scientist as an adviser, the sources say, but was told by senior scientists that he should instead place such a person in a job with real administrative clout. Moniz, who returned to MIT last year after a brief but successful stint at the Office of Science and Technology Policy at the White House, was unavailable for comment on the pending appointment.

Wild boars give clue to Chernobyl fall-out

[PARIS] The French Institute for Nuclear Protection and Safety (IPSN) has confirmed the presence of patches of radioactivity in the Vosges mountains originating from the Chernobyl nuclear accident in 1986.

A helicopter study of a 35-km² area revealed contaminated patches where levels of caesium-137 reached 15,000 becquerels (Bq) per m², three times the national average.

The study was conducted after wild boars were found to be contaminated with 1,500 to 2,000 Bq per kg. The maximum commercial limit is 600. But IPSN believes that the risk to health is low. It argues that hunters eating 200 grams of boar per week would register one-third of the agreed safe levels of radiation.

Company destroys nuclear fire photos

[LONDON] Japan's state-run nuclear power company Donen has admitted destroying photographs taken after a fire at its nuclear fuel reprocessing plant triggered the country's worst nuclear accident.

Donen employees are understood to have taken photographs soon after fire broke out at the plant on the morning of 11 March. An explosion followed some 10 hours later, exposing 37 staff to radiation.

Japan's Science and Technology Agency and the police have so far been unable to reconstruct events after the fire. Police sources told the Kyodo Tokaimura news agency that disposing of the pictures could amount to destruction of evidence.

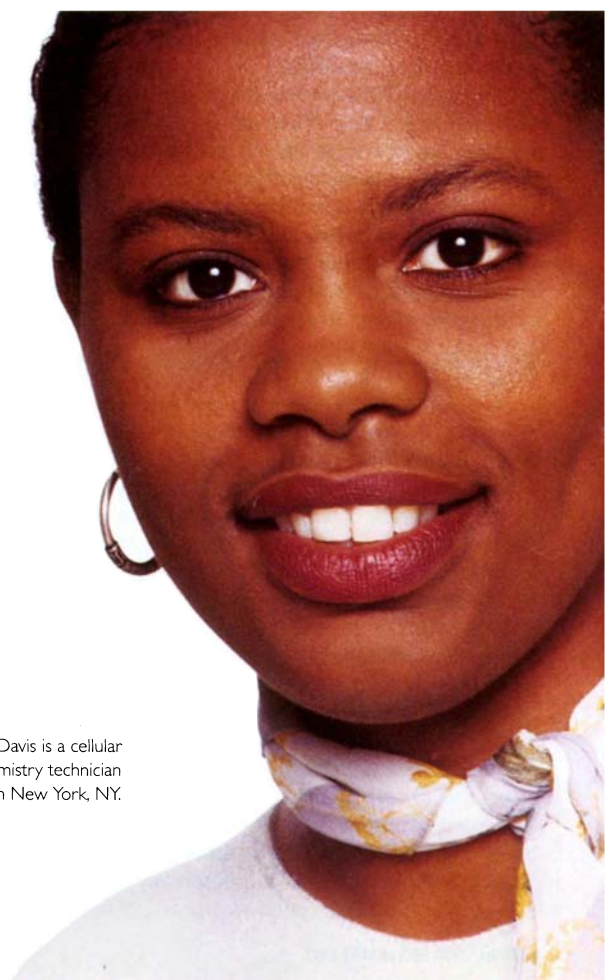
Fossil traders found guilty in exports trial

[SYDNEY] Four men have been convicted for illegally trying to export Australian fossils, in two separate trials in Perth, Western Australia, that ended last month. The first trial, involving David Vaughan and John Bennett, both rock dealers, has set a legal precedent.

The pair are the first Australians to be sentenced, under national heritage laws, for conspiring to export, in this instance, six rare

Obtaining **reliable**
PCR results means
using **Ready-To-Go**
PCR Beads

Terri Davis is a cellular
biochemistry technician
working in New York, NY.



crinoid fossils of Permian age from Western Australia to Germany six years ago.

Both Vaughan and Bennett had pleaded not guilty but Bennett was fined A\$10,000 (US\$7,800). Vaughan's sentence was held over to a second trial with two others for exporting Ediacaran fossils from South Australia. The fossils were later traced to Japan and recovered.

All three pleaded guilty and now await penalties. Vaughan is also accused, alone, in a third case yet to reach court.

NASA in deal with Japan for asteroid mission

[WASHINGTON] The US National Aeronautics and Space Administration (NASA) signed a preliminary agreement last week with Japan's Institute of Space and Astronautical Science (ISAS) to participate in the Japanese MUSES-C mission, which would return the first sample of an asteroid to Earth in 2006. Under the agreement, which has yet to be finalized, NASA would help to test the spacecraft before launch and track it with the giant antennas of the Deep Space Network during its mission. In exchange, NASA would get some of the returned sample material and would include on the spacecraft a 1-kg 'micro-rover' that would drive around the surface of the asteroid Nereus. MUSES-C is scheduled for launch in 2003.

Lecturer faces inquiry over Net comments

[LONDON] Chris Brand, a psychology lecturer at the University of Edinburgh, is facing possible dismissal because of comments published in an Internet newsletter in which he appears to have condoned paedophilia. A university tribunal is meeting this week to decide what action to take.

Brand has been suspended since November after questioning whether paedophilia charges brought against the Nobel-prizewinning scientist Daniel Gadjusek were in the public interest. Brand is understood to believe that non-violent paedophilia between consenting partners over the age of 12 is not harmful providing both have above average intelligence and education.

'Politics grounded space monkeys'

[LONDON] A senior Russian scientist has said he believes that the United States decision not to fly macaque monkeys on board the Russian Bion 12 spacecraft next year was made partly on political grounds in addition to concern for the monkeys' health. Attempts had been made in Congress to scupper the mission (see *Nature* 387, 4; 1997).

However, in an interview with Russia's

Interfax news agency, Yevgeniy Ilyin, deputy director of the Institute for Medical and Biological Studies, said he expected the other planned experiments on birds, rodents, and insects aboard Bion 12 to go ahead. But he warned that new experiments on other animals, in place of the macaques, would require new equipment, and thus additional funding.

Soros Foundation gets Belarus tax bill

[LONDON] The Belarus branch of the Soros Foundation has been told to pay tax on 19 projects. The foundation has said it intends to appeal.

Projects in science, education, culture, health care and environmental protection are exempt from having to pay tax under Belarus government regulations. But following an audit of the Soros Foundation's projects, local tax inspectors decided that a documentary entitled *The History of Small Towns in Belarus*, and a project to create an archive of the Chernobyl nuclear accident, did not fall into tax-exempt categories.

A press statement issued by the foundation said that the tax department seemed to have its own interpretation of the terms 'culture', 'environmental protection' and 'science'.

There are many opportunities for introducing contaminants into your PCR reaction which could jeopardize the reliability of your results. Preparing individual PCR reactions every time you want to perform PCR is not only time consuming, but allows for pipetting errors. Sharing a master PCR mix with others could also compromise the integrity of the PCR reaction, resulting in reduced reliability. Pharmacia Biotech has applied the Ready-To-Go® bead technology representing pre-mixed, pre-formulated, ambient-stable, single-dose reactions to typical PCR reactions, yielding Ready-To-Go PCR Beads.

New ambient-stable Ready-To-Go PCR beads: just add primers, DNA and water

Obtaining reliable PCR results has never been so simple and reliable. You begin by removing a 0.5 ml tube containing an optimized PCR bead from the foil pouch. Next, watch the bead dissolve in seconds as you add DNA, primers, and water. Then, place the tube into a thermocycler and turn it on. That's all it takes to obtain reliable PCR results. There's nothing like it on the market today!

The pre-formulated PCR beads deliver unprecedented reliability and reproducibility for amplifying DNA fragments from a variety of DNA templates, such as plasmid, cDNA, genomic sources and viral DNA. This high level of performance is achieved because every time you perform PCR, a fresh reaction bead is used. Fewer steps means minimizing the risk of contamination, and increasing the reliability of your results. Each batch of Ready-To-Go PCR Beads is function-tested, ensuring lot-to-lot consistency.

To find out more, give us a call: 1 (800) 526-3593 in the USA; +81 3492 6949 in Japan; +46 18 16 50 11 in Europe and the rest of the world. Or visit us on the Internet: <http://www.biotech.pharmacia.se>.

"Purchase of Ready-To-Go® PCR Beads is accompanied by a limited license to use it in the Polymerase Chain Reaction (PCR) process solely for the research and development activities of the purchaser in conjunction with a thermal cycler whose use in the automated performance of the PCR process is covered by the up-front license fee, either by payment to Perkin-Elmer or as purchased, i.e. an authorized thermal cycler.



**Pharmacia
Biotech**

Uppsala, Sweden. (And the rest of the world)



A Ready-To-Go PCR Bead: each is optimized for PCR and contains buffer, nucleotides, and Taq DNA Polymerase. (Magnified 180%).

READER ENQUIRY NO. 655

Mining waste poisons river basin

Sir—The mining of polymetallic ores (tin, zinc, copper, lead, antimony, bismuth, silver, gold, cadmium, arsenic and so on) is the predominant industry of Bolivia. The mining districts of Potosí and Oruro (established in 1545) are located in the eastern margin of the central Andean plateau, the waters and streams of which converge in the great Pilcomayo river. The Bolivian sub-basin of the Pilcomayo covers an area of 98,100 km².

The mining has produced huge quantities of waste materials, including the metals mined, along with mercury from the use of the amalgam method for extracting silver. (Between 1570 and 1900, the total discharge of mercury from silver mining in South America has been estimated at approximately 196,000 tonnes¹.)

In November and December 1996, dead fish with symptoms of heavy metal poisoning, such as red eyes and soft flesh, were found floating in the Pilcomayo waters in western Chaco. These dead fish did not have mud in their gills; the Guaraní and Nivaclé population commonly eat fish which have been asphyxiated by silt. Elementary chemical analyses of the Pilcomayo waters display large amounts of lead, arsenic, mercury and other heavy metals. But the Paraguayan authorities, to protect their international agricultural and cattle exports, have chosen to ignore this pollution. At the same time, a mining dam retaining sulphide metal waste in Bolivia broke, spilling into the Pilcomayo. During the rainy season, the river overflows, changing the composition of the river water, but the heavy metals can be

temporarily sealed in the mud deposits, producing possible future sources of poisoning.

The Chaco is a large tropical plain in the interior of South America, consisting of parts of Argentina, Paraguay and Bolivia (840,000 km² in area). The sediments and material dissolved by the Pilcomayo river flow into the Paraguay river in Asunción, then into the Paraná river (in Corrientes), and the Paraná river finally carries them into the Plata river, in Buenos Aires.

The Guaraní and Nivaclé-Manjui population living in the huge alluvial fan of the Pilcomayo suffer from changes in the course of the river², cholera epidemics (from drinking from polluted watercourses and pools)³ and serious nutritional problems. Furthermore, they drink water directly from the river, and one of their main sources of food is freshwater fish such as Characidae, Pimelodidae, Loricariidae and Curimatidae⁴. The poisoning of the river Pilcomayo with heavy metals is a significant threat to the indigenous population. At the mouth of the Plata river, in Buenos Aires, the situation is better: purifying plants take the clay-rich waters of the river before human consumption, and the heavy metals from the Andean mines have, by this point, spread out into the massive additional flow of the Paraná and the Paraguay. But the poisoned waters of the Pilcomayo cross half the continent and affect millions of people in Argentina, Paraguay and Bolivia.

A few years ago, while carrying out a United Nations assignment in Bolivia⁵, we realized the huge economic difficulties faced by the Bolivian mining industry; all

its resources are directed to establishing joint ventures with foreign companies for the sole purpose of obtaining injections of cash. The mining and environmental authorities do not appear to understand the problems that arise from dumping metallic sulphides in rivers.

This situation is too complex for the Bolivian government to solve on its own and should be looked into by those countries which mine, or have mined, in Bolivia, such as Spain, under the 'Mita' colonial system (before 1801) and Britain, Germany, Japan, Russia and the United States (after 1801) under different paracolonial regimes.

In short, the rich foreign countries that have been involved in the exploitation of Bolivia should take it upon themselves to analyse the pollution of the river Pilcomayo and the stability of the sulphide dumps if the authorities of Bolivia, Paraguay and Argentina continue to turn a blind eye to this complicated problem.

Javier Garcia-Guinea

Museo Nacional de Ciencias Naturales,
Dpto de Geología,
C/ Jose Gutierrez Abascal 2,
28006 Madrid, Spain
e-mail: guinea@fresno.csic.es

Maria Huascar

Reserva Nacional Guaraní y Nivaclé del Pilcomayo,
Departamento de Boquerón,
Mariscal Estigarribia, Paraguay

1. Nriagu, J. *Nature* 363, 589 (1993).

2. Iriondo, M. *Geomorphology* 7(4), 289–303 (1993).

3. Mazzaferro, V. E. et al. *J. Diarrh. Dis. Res.* 13(2), 95–98 (1995).

4. Menni, R. C. et al. *Hydrobiologia* 245(3), 129–146 (1992).

5. Garcia-Guinea, J. *Min. Ind. Int.* 1026, 13–19 (1995).

Explanation for singing sands

Sir—The studies by Goldsack *et al.* of 'singing' sands (*Nature* 386, 29; 1997), proposing that the grains have a silica-gel surface coating, seem to lead in a promising direction. I have long suspected that the source of the sounds, caused in some sands by dragging one's foot across the surface, is some kind of thin coating on the grains. The following observations led me to this notion.

For many years at Duneland Beach near Michigan City, Indiana, at the south end of Lake Michigan, I noticed that dry sands on the berm crest of the beach can generally be made to 'sing', whereas similar sands only metres away on the lower slope of the foredune behind the beach generally cannot. That suggests that the wind action

that sweeps the grains from the beach to the dune must alter them in some way.

That this alteration may be the abrasional removal of a surface coating is suggested by the observation that the beach sand develops a thin fragile crust when it dries out after storm overwash. This frail crust becomes evident above trapped air-bubble holes along the upper part of the swash zone. After drying, the subsurface sand settles to a denser packing, leaving the crust to bridge over the resulting cavities. It may then collapse into what looks like Mars's chaotic terrain when one stamps on the nearby sand. Although Lake Michigan's water is fresh, it must contain sufficient salts to produce a coating on the grains that binds them together and makes them sing when disturbed—but a coating that is easily destroyed by wind abrasion when the grains are blown back into the foredune.

W. N. Melhorn (Purdue University) and

I planned to test this idea by asking S. V. Margolis (University of California, Davis) to look at the grain surfaces under the electron microscope. Margolis had studied other sand-grain surfaces, and we had samples of singing and non-singing sands from a number of sites. But he became preoccupied with another project, and then died of cancer before he could look at our sands.

That ended our study, but the work of Goldsack *et al.* suggests that we were on the right track, and they and others should look at the microtextures of the grain surfaces to see if they may indeed hold a key to the sands' acoustic properties.

Charles G. Higgins*

Department of Geology,
University of California,
Davis, California 95616, USA

*Present address: PO Box 1727, Carmel Valley, California 93924, USA.

The dog did nothing in the night-time

Sir— The recent Briefing on malaria (*Nature* 386, 535–540; 1997) highlights the current critical situation. The accompanying Correspondence (386, 541; 1997), a timely reminder of the need for sustained international collaboration, is signed by individuals representing a broad spectrum of national and international organizations, all supporters and all directly concerned with the problem.

The Scientific and Technological Co-operation with Developing Countries programme (INCO-DC) of the European Commission (EC) Directorate General XII, which was a co-sponsor of the meeting, has for many years supported collaborative North–South research projects in malaria. As European scientists already involved in research networks on malaria supported by INCO-DC, we were surprised at the absence of an EC signatory to the letter. The EC is mandated to embrace complex issues at the forefront of international collaboration, issues likely to be resolved only through the joint efforts of scientists from the developing countries and from the North.

The European Union harbours under-exploited potential, which might be harnessed within the EC, for coordination of the links between development and research funding. This would strengthen both research on malaria and the ability to develop and sustain control programmes within reinforced health services. The increased awareness of the scale of the malaria problem is also timely in the light of current discussions on the content of the Fifth Framework Programme. If Europe really has a serious commitment to solving the problems of malaria, it must seize now the opportunity to play a leading role in the global attack on malaria. It is to be hoped that the spirit of Dakar will be met by a committed and urgent response from the EC.

Alan Thomas, Biomedical Research Centre, Rijswijk, The Netherlands, e-mail: thomas@bprc.nl;
Andrew Waters, University of Leiden, The Netherlands; **Pierre Druilhe**, Institut Pasteur, Paris, France; **Michael Lanzer**, University of Würzburg, Germany; **Mario Coluzzi**, University of Rome 'La Sapienza', Italy

Sir— I should like to reassure your readers that the pharmaceutical industry has not totally given up on new treatments for malaria.

Glaxo Wellcome has recently completed development of a new antimalarial, Malarone, a combination of atovaquone and proguanil, which is active against many

resistant strains of *Plasmodium*. The company has already announced a donation programme for this new drug, which is being piloted in Kenya later this year. Although we believe Malarone will have a significant effect globally on malaria, we recognize that *Plasmodium* has a propensity to develop resistance. So we are actively researching new agents that work by different mechanisms.

Such research activities range from a cooperative screening programme in which Glaxo Wellcome supplies compounds to the World Health Organisation (WHO) for testing against whole parasites to more targeted approaches. Regarding the latter, a number of cDNA sequences for malarial enzymes are appearing in the databases, and lactate dehydrogenase is an example of one such essential enzyme, for which a high-resolution crystal structure has recently been published by Professor J. John Holbrook's group at the University of Bristol (C. R. Dunn *et al.* *Nature Struct. Biol.* 3, 912–915; 1996). Comparative studies with its mammalian counterpart indicate significant differences, making this an attractive target for the rational design of new selective and cidal agents for the treatment of malaria. Thus lactate dehydrogenase forms the focus for an international screening programme involving scientists in Bristol and Glaxo Wellcome sites in Madrid and Stevenage, which we hope will become a model for the future. We have every intention of developing any chemical leads that the screening programme provides and will make use of the excellent cooperation provided by our link with WHO.

Mike S. Marriott

Microbial Diseases Research,
Glaxo Wellcome SpA,
Via A. Fleming 2,
37100 Verona, Italy

Sir— We applaud your articles about the urgent need for new malaria control programmes and for innovative research to guide them.

Your invocation of the legacy of Ronald Ross seems especially apt in this regard, and it reminded us that a critical part of that legacy is his demonstration that the most powerful insights with respect to control derive from research that integrates theory, laboratory and field work. While the advantages of such integrated research programmes may already appear self-evident to readers in the home country of Ross and his illustrious successor, MacDonald, this may be less true elsewhere.

F. Ellis McKenzie

William H. Bossert

Harvard University,
Cambridge, Massachusetts 02138, USA

Cloning humans

Sir— If there were no other reasons to argue against cloning humans, Mike Fainzilber gives at least one very good one (*Nature* 386, 431; 1997).

His use of cloning a child dying in an accident as an example of 'good' cloning suggests a concept of life reduced to running a genetic code. Cloning a dying loved one — just like rerunning an accidentally aborted computer program — deeply questions the dignity and value of human life. Furthermore, the knowledge of being a cloned 'reserve copy' of a dead precursor might result in a deep identity problem for the clone if there is more to personality than just the genes.

Nikolaus Nestle

2N Kreative Entwicklungen,
Holdergasse 10,
D-89291 Holzheim, Germany

Sir— Galibert *et al.* argue for a "total and definitive ban" on human cloning on the grounds that "the essence of humanity lies in the uniqueness of each of its members, resulting from the unique recombination of two equally unique genomes" (*Nature* 386, 431; 1997). What are we to do with the ~1% of the population who are already monozygotic twins (the result of 1 in 250 pregnancies) if they are not essentially human?

Mark R. Baker

University of Oxford,
Radcliffe Infirmary,
Oxford OX2 6HE, UK
e-mail: mark.baker@green.ox.ac.uk

Not quite the same

Sir— Robert Desowitz, in his review of *Deadly Feasts* by Richard Rhodes (*Nature* 386, 565; 1997), describes the author's anger and contempt for various authorities, and says he "serves as a kind of latter-day Sinclair Lewis, who excoriated the Chicago meat-packing industry". Perhaps Desowitz knows that the author of *Main Street*, *Babbitt* and *Arrowsmith* excoriated the meat industry to his friends, but it was Upton Sinclair who did it in print, in *The Jungle* (1906).

Poor Upton Sinclair is always mistaken for his partial namesake; a pity, because he was an interesting man who made a career out of excoriation. In 1923, he began to read from the Bill of Rights at a rally in San Pedro, and was quickly arrested for "discussing thoughts and theories contemptuous of the State of California".

J. J. Cohen

Department of Immunology, B-184,
University of Colorado Medical School,
Denver, Colorado 80262, USA
e-mail: John.Cohen@uchsc.edu

Oil back on the global agenda

A permanent decline in global oil production rate is virtually certain to begin within 20 years. Serious planning is needed to deal with the economic consequences.

Craig Bond Hatfield

In 1985, global oil consumption was 59.7 million barrels per day¹; by 1995 it was more than 69 million barrels per day². This 16 per cent rise in demand was supplied almost entirely by an increase in oil production by members of the Organization of Petroleum Exporting Countries (OPEC) — mostly in the Middle East — from 16.1 to 25 million barrels per day^{1,3}. Growth in production in the North Sea, Latin America and Asia has barely exceeded precipitous declines in US and Russian oil production.

Geological data indicate that, during the next ten years, oil production outside

OPEC countries will remain incapable of significant, sustained growth and is likely to begin a permanent decline during the first decade of the twenty-first century. This seems inevitable because new discoveries of reserves are not keeping pace with production. Globally, new discovery has averaged less than 9 billion barrels per year since 1985 (ref. 4), while consumption has averaged more than 23 billion barrels per year¹. The rate of discovery peaked in the 1960s and has since declined⁵, despite record high rates of exploration in the early 1980s.

Low reserves

Most of the reported 'growth' in oil reserves during the past decade has been from revised estimates: in 1988 and 1989, Venezuela, Iran, Iraq, Abu Dhabi and Saudi Arabia reported upwards revisions totalling 277 billion barrels¹. This accounts for nearly all the growth in global reserves of oil from 1987 (700 billion barrels) to 1990 (1,000 billion barrels). Because it is improbable that these OPEC countries had valid reasons for such gargantuan and simultaneous revisions, it has been suggested⁶ that the increases are political rather than real, perhaps intended to discourage exploration for oil elsewhere in the world or to help establish OPEC production quotas.

If it is the case that the world's oil reserves are indeed as great as 1,000 billion barrels, a continuation of the current oil consumption rate of 26 billion barrels per year would

deplete global reserves by the year 2036. The average estimate of about 550 billion barrels of producible oil yet to be discovered⁷ would add another 21 years in the unlikely event that there is no growth in consumption rate.

But such calculations are unrealistic, because global oil production rate will pass its maximum and begin to decline long before resources are exhausted. For example, oil production in the United States reached its peak in 1970 and has declined since, as was clearly predicted in 1956 (ref. 8). This forecast was the first to apply to oil production a mathematical model that yields growth in production rate until about half of the producible resource has been consumed, after which the production rate declines until the resource is exhausted. It is possible that new oil-recovery technology, as used in the North Sea, could extend growth in production rate beyond the point at which half of the producible resource has been consumed, but in this event the subsequent decline in rate would be accelerated.

How long can growth continue?

For how long can the global oil-production rate continue to grow? Even if the reported reserves were accurate, the addition of the average estimate of about 550 billion barrels of producible oil yet to be discovered would yield 1,550 billion barrels of oil remaining to be produced. Taken together with the approximately 800 billion barrels already consumed, this gives 2,350 billion barrels of 'ultimate' production, which is greater than several recent estimates of ultimate oil production^{6,7}. The mid-point, at which half of the ultimate production will have been consumed, would be the year 2011, if production rate remains at its present level.

But if, as is likely, some of the reported reserves do not exist, or if, as discovery rates of recent years suggest, we have less than 550 billion barrels of producible oil yet to discover, or if consumption continues to grow, cumulative production will reach half of the ultimate production in the first few years of the twenty-first century. In this case, global oil-production rate will peak and begin its decline during the first or second decade of the twenty-first century.

If global demand for oil grows from 1995 to 2005 by 16 per cent, as it did from 1985 to 1995, demand for OPEC oil by the year 2005 will be 36 million barrels a day. OPEC currently produces 25 million barrels a day and has an estimated production capacity of about 29 million barrels a day, excluding Iraqi production. (Iraq's maximum produc-

tion rate over a year is 3 million barrels a day.) So OPEC must increase its production capacity significantly if it is to meet even conservative projections of demand by 2005. What will maximum potential production capacity be between 2010 and 2015, when OPEC will be near (just approaching or just beyond) the mid-point of its ultimate production?

The world probably will reach its maximum oil production rate in the next 15 years. Growth in global oil-consumption rate during 1996 was considerably greater than its average annual growth for the past ten years. Such increases are mainly driven by growing demand in countries where economic growth is accelerating: since 1985, energy use has grown approximately 30 per cent in Latin America, 40 per cent in Africa and 50 per cent in Asia¹.

Energy consumption in developing countries could surpass that in economically developed countries within 20 years. There will be growing competition for a dwindling oil supply, which raises the question of how long standards of living can rise in the developing world and how long they can be maintained in the developed world.

Despite the intensive, intergovernmental debates on the environmental effects of energy policies, geological constraints on the amount of inexpensive fluid fuel that can be produced will soon override governments' decisions about future rates of fossil-fuel burning. The past century of unprecedented economic growth has been based largely on increasing availability of cheap but rapidly dwindling petroleum resources. It is unfashionable in energy policies and economic theory to recognize a time limit on growth in oil consumption rate. But the arithmetic on which my argument is based must be acknowledged. The coming era of permanent decline in oil-production rate and the economic and social implications of this phenomenon demand serious planning by the world's governments. □

Craig Bond Hatfield is in the Department of Geology, University of Toledo, Toledo, Ohio 43606, USA.

1. *Basic Petroleum Data Book: Petroleum Industry Statistics* Vol. XVII, No. 1 (American Petroleum Institute, Washington DC, 1997).

2. *World Oil* 216, 17 (December 1995).

3. Knapp, D. H. *Oil Gas J.* 93, 35–43 (25 December 1995).

4. Laherrere, J. *World Oil* 215, 33 (January 1994).

5. Ivanhoe, L. F. *World Oil* 216, 77–88 (October 1995).

6. Campbell, C. J. *The Golden Century of Oil 1950–2050* (Kluwer, Dordrecht, 1991).

7. Masters, C. D., Root, D. H. & Attanasi, E. D. *Science* 253, 146–152 (1991).

8. Hubbert, M. K. in *Drilling and Production Practice*, 7–25 (American Petroleum Institute, Washington DC, 1956).

Down or out in blood and lymph?

Simon Wain-Hobson

New studies indicate that relatively few infected cells can sustain an HIV infection. Combining this knowledge with powerful multi-drug therapy, it may now be possible to eradicate HIV in a patient within two to three years. Time will tell.

With black humour the microbiologist asks, "What's the difference between true love and a herpesvirus infection? Herpes is forever". Like it or not, this is a brutal reminder of the difficulties of ridding the body of some of its most abiding visitors. Re-activation of the childhood chickenpox agent (varicella-zoster virus) in later life as shingles is but one example. Retroviruses also persist — which brings us back, once again, to human immunodeficiency virus (HIV).

Two papers in this issue illustrate different facets of infection by HIV. On page 183, Chun *et al.*¹ examine the lymph nodes of infected patients, and conclude that HIV infection is sustained by relatively few infected cells. Perelson *et al.*² are brave enough, on page 188, to use the words therapy and eradication in the same abstract. They studied HIV-infected patients after treatment with potential antiviral agents, and they believe that it may be possible to rid patients of HIV within two to three years.

HIV encodes a number of enzymes, notably a protease (protein-digesting) and a reverse transcriptase, the enzyme that metamorphoses RNA into DNA. Enzymes are good targets for drug development, so by attacking both with combinations of such drugs, remarkable advances have been made in the treatment of HIV infection. But the situation is far from perfect. New inhibitors are constantly coming onto the scene, and Perelson *et al.*² now report the results of a study of nelfinavir — a newly developed protease inhibitor — in combination with two extant drugs effective against the viral reverse transcriptase.

Within ten days or so, Perelson *et al.* found that the levels of HIV in the blood plasma of eight infected patients could be driven down by two orders of magnitude or more, with a half-life of around 0.6 days³⁻⁵. Thereafter, a second-phase decline in the viral load, albeit with a much longer half-life (6–25 days), was observed. Using elegant mathematical modelling, which is now commonplace in HIV research, the authors explored the possible meaning of the two phases.

For our story, the point is that the number of infectious virus particles (virions) was still going down after more than 40 days with this

particular triple combination therapy. Extrapolating from this second phase, Perelson *et al.*² conclude that, barring drug resistance or long-term complications of therapy, it should be possible to eradicate HIV in a patient within two to three years.

Only time will tell whether, with this particular regimen, there may be a third phase of viral infection with an even longer half-life. Or perhaps there are still infected cells in some sanctuary, far from the reach of drugs or the immune system? A blunt personal view is never to think that you've outsmarted HIV; it should be obvious if, and when, eradication is achieved. As it is, triple therapy is hugely expensive, so in an increasingly ungenerous world success may well be bitter-sweet.

In the other paper, Chun *et al.*¹ take us into the world of infected lymph nodes. Here, foreign molecules (antigens) encounter the T, B and antigen-presenting cells of the immune system. Many cells of the immune system have molecules on their surface that often betray the identity of the cell. The CD4 molecule is one such protein (Fig. 1). CD4 is found on (and defines) a helper group of T cells that are essential for efficient

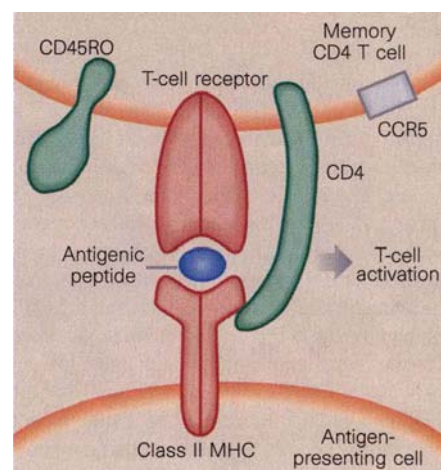


Figure 1 A mêlée of membrane molecules. Foreign peptide is presented by two molecules — the class II major histocompatibility complex (MHC) on the surface of an antigen-presenting cell. This complex is recognized through a pair of T-cell-receptor molecules on an antigen-specific CD4-positive (memory) T cell. This process is abetted by an interaction between CD4 and one of the MHC class II molecules, and T-cell activation ensues. To get through the CD4-positive T-cell membrane, HIV hijacks a pair of surface molecules, CD4 and CCR5.

coordination of immune responses. By virtue of their antigen-specific surface receptors, they may sift through, and identify, the plethora of antigens. As non-dividing, resting cells, they search for antigen, upon encounter with which they become activated and start to divide. HIV uses CD4 to gain entry into these important helper T cells, and loss of CD4-positive cells is the defining immunological trait of AIDS. So Chun *et al.* have asked, how many resting and activated

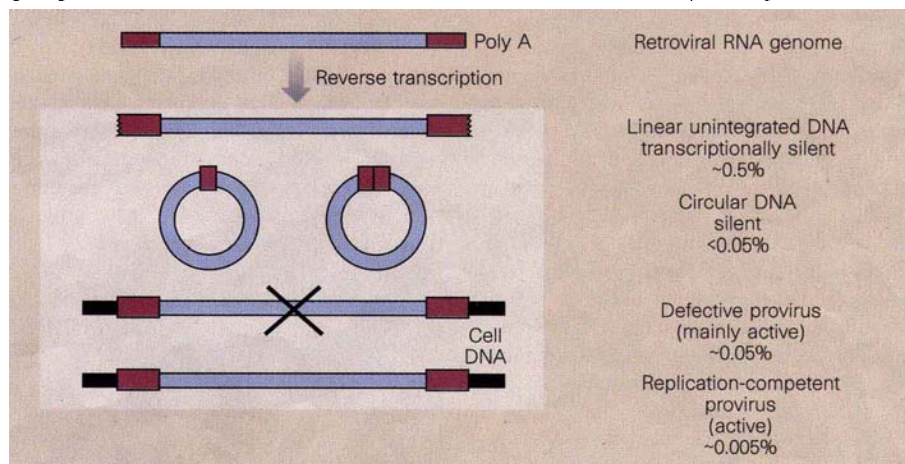


Figure 2 Form, function and frequencies of HIV DNA. HIV encapsulates an RNA genome that has redundant (non-coding) sequences at either end. Once HIV is injected into a host cell, it uses reverse transcriptase to convert its genome into a double-stranded DNA molecule with rearranged ends. The linear genome may then be integrated randomly into the host-cell chromosome, giving rise to the provirus. But errors in reverse transcription mean that some of these proviruses are defective (that is, they cannot produce infectious virus), although viral proteins may be produced. Chun *et al.*¹ have found that most of the HIV DNA is in an unintegrated linear form with nibbled ends, and a smaller proportion exists as ligated circles. They have also measured the frequency of resting CD4-positive cells that harbour the various forms of HIV.

CD4-positive T cells harbour HIV DNA, and in what form?

The familiar model has it that reverse transcription of the HIV RNA retroviral genome results in double-stranded DNA that is subsequently integrated into the host-cell genome. Only this form — the 'provirus' — can support efficient transcription. Inevitably, perhaps, there are variations on this theme — the genome may also hang around as a linear unintegrated form, or be circularly ligated upon itself. Because reverse transcription is an error-prone process, it may lead to the introduction of inactivating mutations and, as if that wasn't enough, HIV can remain transcriptionally silent (or latent) in resting CD4-positive T cells. In this case, HIV only goes into overdrive when the T cell is activated by antigen.

All of these factors complicate the analysis of HIV in lymphoid organs. Undaunted, Chun *et al.* applied a battery of techniques based on the polymerase chain reaction, which allowed them to dissect out the different HIV forms (Fig. 2). They carried out lymph-node biopsies, and found that although about 0.5% of resting CD4-positive T cells harbour HIV DNA, only around 0.05% carry proviruses. Moreover, of these, only 10% may produce HIV virions. These seem to be small proportions, but they translate into large absolute numbers — about 10^9 HIV-containing CD4-positive T cells, of which about 1.2×10^7 contain latent provirus, and 1.4×10^6 may produce infectious virus (the range is ± 0.5 logs). Chun *et al.* also looked at the total number of CD4-positive T cells (that is, activated and resting) containing HIV, and they found values two- to threefold greater at most.

The inescapable conclusion is that HIV infection is sustained by relatively few infected cells. Yet viruses invariably pack a big punch, and up to 10^2 – 10^4 virions may be produced per cell. HIV is no different⁶ and, accordingly, the dynamics of HIV production — between 10^8 and 10^{10} virions per day, depending on the disease stage^{3–5} — may be readily understood. Surprisingly, however, the relative proportions of the HIV DNA forms were comparable whether the CD4 cells were recovered from blood or lymph. But the fact that the freeway (blood) reflects quite well the workings of the inner city (lymph nodes) considerably simplifies longitudinal studies of HIV dynamics.

Despite this, the connection between longitudinal studies of viral dynamics through the analysis of blood, and cross-sectional analyses of HIV in lymph-node biopsies is incomplete. Perelson *et al.*² show that the decay of peripheral blood cells that can produce HIV parallels that of the second-phase decay of viral levels in blood plasma. Their modelling indicates that the virus in the second phase of decay comes from long-lived infected cells, perhaps macrophages (scav-

enging white blood cells). Yet Chun *et al.*¹ identified a total body load of some 1.4×10^6 replication-competent proviruses in resting CD4-positive T cells. More precisely, they studied the CD45RO-positive cell population, which corresponds roughly to the memory-cell population (Fig. 1). Immunological memory and dynamics⁷ are very complicated, with many compartments and controversies. Be that as it may, the CD45RO-positive cell population turns over with a half-life of some 22 weeks⁸. Does this presage a third phase of viral decay?

How do these findings relate to the loss of CD4-positive T cells? From the slope of a graph of CD4-positive cell decline, a net loss of some tens of millions per day can be calculated — a figure that is larger than the total number of productively infected CD4-positive T cells. Does this mean that some of the other HIV DNA forms contribute to pathogenesis? For example, cells containing apparently defective proviruses are about tenfold

more abundant than the productively infected cells. Although these cells cannot produce virus, if some viral proteins were produced, would they mark out the cell as being foreign, so bringing down upon it the wrath of HIV-specific immune response⁹? The bottom line is that nobody will be satisfied until there is an adequate explanation for the loss of CD4-positive T cells. Despite this quandary, HIV can clearly be put down by therapy. So might it even bow out? Cross your fingers. □

Simon Wain-Hobson is at the Institut Pasteur, 28 Rue du Dr Roux, 75724 Paris cedex 15, France.

1. Chun, T.-W. *et al.* *Nature* **387**, 183–188 (1997).
2. Perelson, A. S. *et al.* *Nature* **387**, 188–191 (1997).
3. Ho, D. D. *et al.* *Nature* **373**, 123–126 (1995).
4. Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M. & Ho, D. D. *Science* **271**, 1582–1586 (1996).
5. Wei, X. *et al.* *Nature* **373**, 117–122 (1995).
6. Dimitrov, D. S. *et al.* *J. Virol.* **67**, 2182–2190 (1993).
7. Freitas, A. A. & Rocha, B. B. *Immunol. Today* **14**, 25–29 (1993).
8. McLean, A. R. & Michie, C. A. *Proc. Natl Acad. Sci. USA* **92**, 3707–3711 (1995).
9. Wain-Hobson, S. *Nature* **366**, 22 (1993).

Galactic structure

Strange colours peek from dark halo

Penny D. Sackett

Astronomers argue about dark matter. How much of it is there? Where is it? What is it made of? But they agree that the dark matter that makes up most of the mass of spiral galaxies is *dark*, and has a more extended distribution than that of the luminous stars and gas. That is why the discovery three years ago of a faint luminous halo around the nearby, edge-on spiral NGC5907 (Fig. 1), with the same spatial distribution expected for the galaxy's dark matter¹, was met with interest, surprise and scepticism. The critical question was whether the faint light was being emitted by the 'dark' matter itself or by stars. Had light from dark matter been detected or not? The new observations reported by Rudy *et al.* on page 159 of this issue² sharpen rather than answer this question, by implying that a very peculiar population of stars may be responsible for the halo light.

Some sort of dark matter is needed to

explain the rapid rotation of spiral galaxies and the relative velocities of galaxies in clusters, and to tidy up cosmological theories. It needn't all be the same stuff; conventional cosmology requires that most of it is non-baryonic, that is, not made from protons and neutrons. But in galaxies, dark matter may be ordinary baryonic material in a non-luminous form — fading stellar remnants, comets, or brown dwarfs and faint stars. The new results from NGC5907 support — but do not prove — the last of these possibilities.

The precise amount of infrared (IR) light coming from far above the galactic disk of NGC5907 is rather uncertain³, but its distribution appears to be roughly consistent with that of visible light^{1,3}. What makes this faint halo unlike previously detected stellar halos, but similar to the spatial distribution of dark mass, is that the decline of emission with increasing height above the plane^{1–3} is so gradual.



Figure 1 The spiral galaxy NGC5907. Its faint halo, invisible here, appears to be made of an inexplicable mixture of stars.

The new, but confusing, clue about the nature of the faint emission from NGC5907 has been obtained by comparing the spectral energy distribution of the halo — the proportion of light emitted at different wavelengths — to theoretical stellar populations with a given chemical composition and mix of light and heavy stars. Broadly speaking, stars of low mass and those with compositions rich in the elements heavier than helium ('metal-rich' stars) have cooler atmospheres than their high-mass or metal-poor counterparts, and thus emit more of their energy in 'redder' or longer-wavelength regions of the spectrum. Because low-mass stars are more numerous and emit less light per unit mass than high-mass stars, they have been considered plausible dark-matter candidates in the past.

The best match to the new data of the models tested by Rudy and collaborators is a stellar mix that is overwhelmingly dominated by low-mass stars (every 10-fold decrease in stellar mass resulting in a 30,000-fold increase in numbers) with the same metal-richness as the Sun. This conclusion is both tantalizing and disturbing. Tantalizing because such a population could indeed be responsible for all of the dark mass in NGC5907. Disturbing for two reasons. First, stars in the halo of our own Galaxy are much more metal-poor than the Sun, presumably because they formed before most heavy elements were released from stellar interiors in the later stages of stellar evolution. Second, recent Hubble Space Telescope observations sensitive enough to count the numbers of even very faint low-mass stars in the halo of our own Galaxy have failed to find them⁴⁻⁷ in the numbers being inferred for NGC5907. If the spectral energy distribution of NGC5907 has been properly measured and interpreted, it would seem that the dark-matter halos of our Galaxy and NGC5907 are composed of entirely different constituents, even though the two galaxies are similar in many other ways — a conclusion that is, to put it mildly, uncomfortable.

But what other possibilities remain open for the origin of the observed faint halo emission from NGC5907? Could the light be emitted from an ordinary stellar halo, which is faint but does not contain much mass? Probably not. At all wavelengths the spatial distribution is considerably more extended than that of typical stellar halos. Furthermore, the optical colours indicate a population nearly devoid of giant stars³, and combined with the infrared colours suggest a population that is too metal-rich and far too highly dominated by low-mass stars to be typical².

Could the light be from the stellar debris of a small satellite galaxy that has been sheared apart and accreted onto NGC5907 through a gravitational merger, like those sometimes observed for other galaxies,

including our own? Unlikely. That might account for the extended spatial profile of faint light, the asymmetry seen in some observations, and perhaps the presence of metal-rich stars in a halo environment. But the colours are extremely odd, and if the stellar mix is as Rudy and collaborators suggest, the 'satellite' would have been much more massive than NGC5907!

Could the interpretation of the peculiar halo colours for NGC5907 be wrong? Possibly. Calculating the colours of metal-rich, low-mass stars is complicated by the presence of many molecular absorbers in the cool, outer stellar atmospheres. Recently refined models, however, were used by Rudy and collaborators. A critical step will be to verify the prediction that the IR emission is being generated by a large number of dim dwarf stars rather than a small number of bright giants.

Could the measurements themselves be in error? For a single observation, possibly. The faint emission levels mean that detector response and backgrounds must be calibrated precisely⁸. But the unusual colours of the NGC5907 halo rest on several independent measurements^{1-3,9}. A conventional explanation would require that more than one of these measurements be wrong.

This is not the first time in recent months that an unconventional stellar population has

been proposed for halo dark matter. Based on its first two years of microlensing results, the MACHO team has suggested that 50 per cent of the Milky Way's dark matter may be in the form of white dwarfs¹⁰ — compact, faint end-products of stellar evolution, each about half the Sun's mass. But they point out that their results are also consistent with our Galaxy's dark mass being mostly stars and brown dwarfs in the 0.1-solar-mass range. Both possibilities require extremely peculiar mixes of faint objects too dim to be detected by current direct searches in the Milky Way.

So, where does all this leave astronomers? Arguing about dark matter. □

Penny D. Sackett is in the Kapteyn Instituut, Universiteit Groningen, Postbus 800, 9700 AV Groningen, The Netherlands.

1. Sackett, P. D., Morrison, H. L., Harding, P. & Boroson, T. A. *Nature* **370**, 441–443 (1994).
2. Rudy, R. J., Woodward, C. E., Hodge, T., Fairfield, S. W. & Harker, D. E. *Nature* **387**, 159–161 (1997).
3. Lequeux, J., Fort, B., Dantel-Fort, M., Cuillandre, J.-C. & Mellier, Y. *Astron. Astrophys.* **296**, L13–L16 (1995).
4. Bahcall, J. N., Flynn, C., Gould, A. & Kirhakos, S. *Astrophys. J.* **435**, L51–L54 (1994).
5. Flynn, C., Gould, A. & Bahcall, J. N. *Astrophys. J.* **466**, L55–L58 (1996).
6. de Marchi, G. & Paresce, F. *Astrophys. J.* **476**, L19–L22 (1997).
7. Graff, C. & Freese, K. *Astrophys. J.* **456**, L49–L52 (1996).
8. Morrison, H. L., Boroson, T. A. & Harding, P. *Astron. J.* **108**, 1191–1207 (1994).
9. James, P. & Casali, M. M. *Spectrum* No.9, March, 14–16 (1995).
10. Alcock, C. *et al. Astrophys. J.* (submitted).

Evolutionary biology

A chilling tale from the ends of the Earth

As summer time approaches and temperatures soar, the warming seas entice many people to take a dip. But spare a thought for the fish that live in the waters of the Antarctic and Arctic, which can reach a low of -1.9°C . As well as finding food in the frigid polar oceans, these fish must ensure that they do not, literally, freeze.

Fortunately, they have special blood-serum 'antifreeze' glycoproteins (AFGPs) that deter the growth of ice crystals and lower the freezing temperature of the fish

to below that of the surrounding sea. Now, in the 15 April issue of *Proceedings of the National Academy of Sciences* (Chen, L. *et al. Proc. Natl Acad. Sci. USA* **94**, 3811–3822; 1997), Liangbiao Chen and colleagues report that they have found that fish from opposite ends of the Earth have very similar AFGPs but, incredibly, the genes for these antifreeze proteins evolved through completely different pathways.

Members of the Antarctic notothenioid family of fish (pictured) have an AFGP that is made up of a simple sequence of repeating tripeptide units, and it is thought to have evolved from the pancreatic enzyme trypsinogen. Arctic cod also have a tripeptide-repeat-based AFGP, but this shares no sequence similarity with the trypsinogen gene. Moreover, morphological and palaeoclimatic findings indicate that the AFGPs from the two types of fish could not have evolved from the same ancestor. So they must have evolved their respective AFGPs separately and, through convergent evolution, arrived at the same sequence. A chilling tale indeed.

Allison Mitchell

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

They might be giants

John Kappelman

Body mass is a critical determinant of many aspects of an organism's physiology, adaptations and behaviour. Because body mass can easily be quantified for living forms, and because it correlates strongly with life-history parameters such as metabolic rate, longevity and gestation length, much research has gone into one intriguing question — why is an organism the size that it is?

To address the larger issue of how an organism came to be the size that it is, we need to add data from the fossil record, and nowhere has this seen more attention than in palaeo-anthropology. But how large were fossil humans? This question has often been asked for the

australopithecines (dating from four to one million years before present; Myr BP) and now, on page 173 of this issue, Ruff *et al.*¹ describe a new and potentially robust approach to resolving the debate about our genus, *Homo*.

At first glance, calculating body mass should be fairly straightforward because there is such a vast data set from living humans on which to draw. Forensic science offers many regression equations (some of which are restricted by sex and ethnic group²) for accurately estimating stature and body mass from even fragmentary bones. Modern human body form also shows adaptations to climate. Our species is found from the Equator to the Poles and, although our cultural trappings offer some insulation from these climatic extremes, modern human body form has nonetheless adapted according to the rules of surface-area-to-volume scaling: for example, people from more northerly latitudes are usually taller and broader than equatorial folk³. All of these factors must be considered when estimating the body mass of extinct humans.

Ruff *et al.*¹ have examined the record of fossil humans attributed to the genus *Homo* across the past 2 Myr. They divide the sample by low (<30°) and high latitude, and use various regression equations to estimate body mass for individual fossil specimens. The authors suggest that archaic *Homo* (dating here from nearly 2 Myr to 36,000 yr BP) was, on average, about 13% larger than the living worldwide human average. Body mass then seems to have increased through time, with late archaic *Homo sapiens* — dating from just 75,000–36,000 yr BP and often collectively called 'Neanderthals' (Fig. 1) — being about 30% larger than the living worldwide human average (or about 24% larger than living high-latitude humans). But modern humans (dating in this study from 90,000 yr to the present, and overlapping in time with archaic *Homo sapiens*) show a decrease in body mass.

The fossil *Homo* sample from the past two million years is one of the largest ever to be included in a study of this type. But, even here, the issue of incompleteness in the fossil record rears its ugly head: 87% of the fossil humans are from the past 200,000 yr, or just 10% of the total time sampled; and 43% are from the past 20,000 yr, or just 1% of the total time sampled. The sample is heavily skewed towards Africa and Europe (82%), with some temporal groups biased by geographical region and sex. Although new dates⁴ and future discoveries from the oldest part of the

record and other parts of the world could alter these general conclusions, the pattern of an increase in body mass through time is probably quite robust.

Once good body-mass estimates are to hand, the attraction of examining the pattern of brain evolution is irresistible. Brain size is only meaningful when judged relative to body mass, but cases where skulls (which provide the estimate of brain size) are found with their associated skeletons (which traditionally offer the estimate of body mass) are rare. In fact, for the sample included here, only 11% of the specimens have the necessary associated skull and skeleton and, of these, over 75% are modern humans — a statistic that attests to our species' preoccupation with burial of the dead. There are only two associated specimens that are older than 100,000 yr, and these are separated from each other by nearly one-and-a-half million years.

Ruff *et al.*¹ address this problem by dividing the sample into nine time periods, and comparing the relative brain-size estimate (here an encephalization quotient) for the much smaller samples of associated skeletons and skulls with those for the unassociated specimens. The fact that the two values are similar in those cases where the comparison can, in fact, be made, indicates that this approach may be robust. But I worry about the consequences of combining heads from Asia with bodies from Europe, especially because these might be from different species.

The authors conclude that there is a long period of stasis in the evolution of relative brain size, from 1.8 Myr to between 600,000 and 150,000 yr BP, with the main increase starting sometime within this poorly sampled latter interval. The issue of stasis in the early record of human brain evolution has long been debated^{5,6}, although it is supported by the pattern of absolute brain-size through time⁷. This new work generally agrees with an earlier study based on cranial remains only, which used the area of the orbital aperture from a sample of apes and humans as an indicator of body mass⁸.

A possible drawback to using regression equations that are derived from our species only is that the fossil ends up being modelled like a living human with skeletal elements of that size or a body of those proportions. Other data show that archaic *Homo* had a more strongly constructed skeleton than all but the very earliest modern humans⁹, and the pronounced muscle markings on the bones are believed to indicate great strength¹⁰. Along with the observation that modern humans have decreased in body mass, this evidence leads me to wonder whether average living humans offer the best model for archaic *Homo*? Perhaps world-class athletes — at least those before the widespread use of anabolic steroids — offer a closer approximation to the body build and

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Figure 1 Early heavyweights — Neanderthals (75,000–36,000 yr ago) were about 30% larger than modern-day humans.

mass of archaic *Homo sapiens*. Tanner¹¹ reported on the mass and proportions of athletes in the 1958 British Empire and Commonwealth Games in Cardiff, and the 1960 Olympic Games in Rome. The competitors who most closely approximate the build of Neanderthals were the throwers, weight-lifters and wrestlers. Some of these men weighed as much as 91 kg, even though they were narrower across the hips than most Neanderthals. Using a larger, more muscular living human reference sample could produce even larger and perhaps more realistic body-mass estimates.

The pattern of body build and mass that is emerging for archaic *Homo* will require critical re-thinking about the palaeobiology of Pleistocene *Homo*. Although detailed reconstructions of life-history parameters based on estimates of body mass alone are fraught with difficulty¹², the behaviour of archaic *Homo*, with their larger and presumably more muscular bodies, relatively smaller brains, and simple tools, was probably decidedly non-modern — and more dependent on brawn than brains.

So what selection pressures maintained such large body mass with virtually no increase in relative brain size across at least one-and-a-half million years? The movement to higher latitudes is certainly part of the answer, but some of the specimens from Kabwe in Zambia¹, as well as the specimen from Berg Aukas in Namibia, which is estimated at 93 kg (ref. 13), indicate the existence of very large individuals at low latitudes. Large male body mass within our ape relatives seems to be, at least in part, a consequence of competition among males for access to females, and perhaps this sort of sexual selection operated during the Pleistocene (around 2 Myr ago).

The relatively rapid decrease in modern human body mass during the past 90,000 yr is a dramatic contrast to the large body mass of archaic *Homo* during the preceding two million years. What selection pressures could have resulted in both smaller body mass and larger relative brain size in modern humans? These changes do not seem to be tightly linked to technological innovation, although the less sturdily constructed skeleton implicates different behaviours, suggesting that modern humans adopted increasingly less active lifestyles¹⁴.

Now, rather than focusing solely on models that favour selection for ever-larger brains, we should examine the possibility that the pattern in modern humans was driven by selection for smaller bodies, perhaps favoured by a social structure that relied on more cooperative foraging and better communication skills⁸. The high-energy-requirements of the large brain¹⁵ could have been satisfied if this behaviour produced a more predictable and higher-energy diet. The increase in relative brain size of

modern humans may then be, in part, an effect of selection for smaller body mass⁸. Whatever the final answers, this topic will certainly see a flurry of activity as we wrestle with the biological issues of our larger cousins from the Pleistocene. □

John Kappelman is in the Department of Anthropology, The University of Texas, Austin, Texas 78712-1086, USA.

1. Ruff, C. B., Trinkaus, E. & Holliday, T. W. *Nature* **387**, 173–176 (1997).
2. Trotter, M. *Am. J. Phys. Anthropol.* **10**, 463–514 (1952).
3. Ruff, C. B. *Yb. Phys. Anthropol.* **37**, 65–107 (1994).
4. Swisher, C. C. *et al. Science* **274**, 1870–1874 (1996).

5. Rightmire, G. R. *Paleobiology* **7**, 241–246 (1981).
6. Wolpoff, M. H. *Paleobiology* **10**, 389–406 (1984).
7. Leigh, S. R. *Am. J. Phys. Anthropol.* **87**, 1–13 (1992).
8. Kappelman, J. *J. Hum. Evol.* **30**, 243–276 (1996).
9. Ruff, C. B., Trinkaus, E., Walker, A. & Larsen, C. *Am. J. Phys. Anthropol.* **91**, 21–53 (1993).
10. Trinkaus, E. *The Shanidar Neanderthals* (Academic, New York, 1983).
11. Tanner, J. M. *The Physique of the Olympic Athlete* (Allen & Unwin, London, 1964).
12. Smith, R. J. *Curr. Anthropol.* **37**, 451–481 (1996).
13. Grine, F. E., Jungers, W. L., Tobias, P. V. & Pearson, O. M. *Am. J. Phys. Anthropol.* **97**, 151–185 (1995).
14. Lieberman, D. E. & Shea, J. J. *Am. Anthropol.* **96**, 300–322 (1994).
15. Martin, R. D. *Nature* **293**, 57–60 (1981).

Atomic physics

Angling for an anapole

Patrick Sandars and Kendrick Zetie

It is not every day that an entirely new property of the atomic nucleus is observed. But that is what a very precise laser experiment¹, reported by the Wieman group from the University of Colorado, has done. They have measured atomic parity violation in caesium, and found a small difference in the effect for the two transitions that they observe. The most likely explanation for this is the existence on the atomic nucleus of a long-predicted, but never previously observed, parity-violating magnetic moment — the anapole moment.

It is elementary physics that circulating currents give rise to a dipole magnetic moment, with its associated field (Fig. 1a). The dipole field is long-range, falling off as $1/r^3$. The anapole moment, on the other hand, arises from a toroidal current distribution (Fig. 1b), and the field is 'local', in that it does not extend beyond the current distribution. If these two current distributions are added together, one obtains a spiral current, which has a handedness, or chirality. That is, the pattern clearly reverses its handedness

when reflected in a mirror. Chirality in an atom or nucleus implies the violation of parity (mirror symmetry).

We have known since 1957 that the weak force violates parity²: in beta decay, the emitted electrons are more commonly left-handed than right-handed*. Very soon afterwards, Zel'dovich³ predicted that such an interaction could give rise to anapole moments. Although the fundamental form of the weak force is now well known, calculating the anapole moment of a nucleus is difficult because of the perturbing effect of the strong force (the binding force within nuclei). A number of calculations have been made, with a spread of a factor of 2.5 in the predicted value.

Because their magnetic fields are internal, anapoles are not easy to see experimentally; one needs to penetrate the internal structure of the particle. But as nuclear-spin-dependent structure in atomic spectra (hyperfine structure) demonstrates, the

*One of the discoverers of parity violation, Chien-Shiung Wu, died in February. Her obituary appeared in *Nature* **386**, 334 (1997).

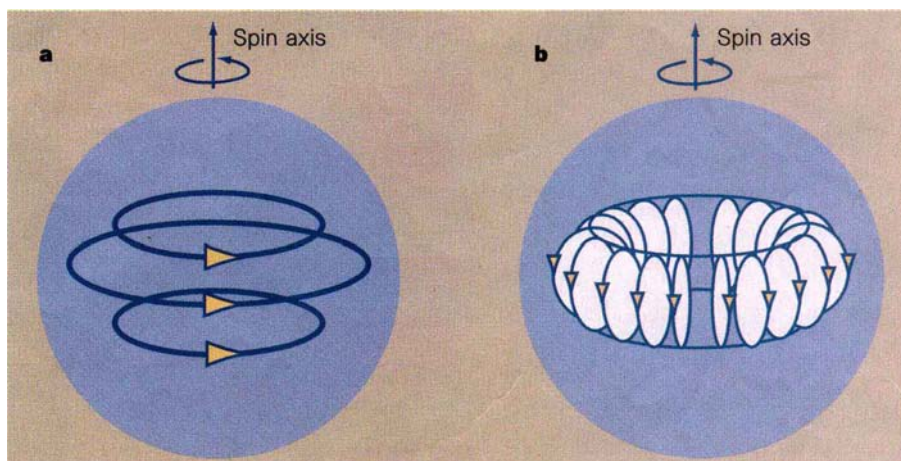


Figure 1 a, Circulating currents give rise to a magnetic dipole moment. By symmetry, the magnetic moment of an isolated particle such as an atom or molecule must point in the same direction as its total angular momentum. b, A toroidal current produces an anapole moment. It is again aligned with the symmetry axis, but it reverses relative to the overall angular momentum when reflected in a mirror. This parity-violating current flow has now been detected in the nucleus of caesium.

electrons do spend time inside the nucleus. So the electron's magnetic moment should interact with the internal anapole fields: the energy of interaction between the anapole, A_N , and the electron current j_e , is of the form $A_N \cdot j_e$.

Because the anapole lies along the nuclear spin, this interaction gives rise to atomic parity nonconservation that depends on the hyperfine state. In contrast the main parity-violating effect in atoms, where exchange of Z bosons changes the electron wavefunctions, is independent of nuclear spin. The relative magnitude of the two effects is easily estimated, independently of detailed atomic theory. With the predicted values of A_N for the caesium nucleus, one expects the anapole effect to be about one per cent of the size of the main effect.

Parity nonconservation is measured in the Boulder experiments^{1,4} in a normally forbidden transition, the 6S–7S line at 540 nm. Applying an electric field, however, creates a known transition amplitude, which interferes with the amplitude induced by the parity-violating interaction. This causes the transition rate to depend on the electric and magnetic fields and the polarization of the light in a way that reverses in a mirror; formally, it depends on $\mathbf{E} \times \mathbf{B}$, where $\mathbf{E}$ is the polarization of the exciting radiation, $\mathbf{E}$ is the electric field and $\mathbf{B}$ is the magnetic field. The presence of the nuclear anapole moment is indicated by a small difference in the parity-violating effect as measured between transitions involving different hyperfine levels. The fractional difference was about 4.7 ± 0.7 per cent, at the high end of theoretical predictions.

The current apparatus is essentially an upgraded version of Wieman's earlier experiment⁴. Improving the accuracy was not simply a matter of increasing the signal-to-noise ratio, although that was a vital first step. The limiting factor in parity-violation experiments has traditionally been systematic errors, effects that can mimic the tiny signal and lead to a false result. So most of the effort in this remarkable experiment has been in understanding and eliminating systematics.

The caesium source is an oven with an aperture formed of an array of capillaries, allowing a wide, collimated beam of atoms to escape. The atoms are optically pumped using two diode lasers to prepare them in a hyperfine substate. The caesium beam then passes through a region of very stable magnetic and electric fields where the 6S–7S transition is excited by an elliptically polarized laser beam at 540 nm. The atoms then pass downstream into the detection region, where those of them that underwent the 540-nm transition are excited by a third diode laser and made to fluoresce. The strength of this fluorescence depends on the parity-violating effect.

The increase in signal-to-noise ratio over their previous experiment comes from three major changes: the use of a spin-polarized beam, improved detection efficiency, and a power build-up cavity for the 540-nm light. This cavity produces about 2.5 kilowatts of laser power, held in a standing wave.

Experimental runs added up to more than 350 hours, but measuring and testing for systematic errors took much longer, and it has taken 12 years of work to achieve such high accuracy. Many of the tests were designed to eliminate known possible systematics, but it was also important to vary

experimental parameters that were *not* expected to affect the result. With many such 'redundant' checks, the group was finally satisfied that there were no errors "lurking in the darkness", and their confidence seems well justified. □

Patrick Sandars and Kendrick Zetie are at the Clarendon Laboratory, University of Oxford, OX1 3PU, UK.

1. Wood, C. S. *et al. Science* 275, 1759–1763 (1997).
2. Wu, C. S., Ambler, E., Hayward, R. W., Hopper, D. D. & Hudson, R. P. *Phys. Rev.* 105, 1413–1415 (1957).
3. Zel'dovich, Y. B., *Sov. Phys. JETP* 33, 1531–1533 (1958).
4. Noecker, M. C., Masterson, P. B. P. & Wieman, C. E. *Phys. Rev. Lett.* 61, 310–313 (1988).

Language impairment

Listening out for subtle deficits

D. V. M. Bishop

"In general, language acquisition is a stubbornly robust process; from what we can tell there is virtually no way to prevent it from happening short of raising a child in a barrel." Steven Pinker's provocative claim¹ has raised many a hackle, but it has a core of truth — language learning is a remarkably resilient human characteristic and, within four short years, most children learn to communicate fluently and efficiently. Because grammatical development is relatively impervious to environmental influences, Noam Chomsky² proposed that language should be regarded as a genetically determined 'mental organ' that develops according to a biological program, just like the lungs, liver or heart.

The paper by Wright *et al.*³ on page 176 of this issue is concerned with children who have a 'specific language impairment' (SLI). These children are of particular theoretical interest because they don't follow the usual smooth course of language acquisition. The language of affected children lags far behind other areas of development for no apparent reason, and they show normal intelligence in tests that require no language⁴. The precise nature of the language difficulty can vary from child to child, but grammar and phonology (that is, production of speech sounds) are typically the most obvious areas of difficulty. For example, a 6-year-old's rendering of the sentence, "Goldilocks ran away from the three bears because she thought they might chase her" may sound more like, "Doediloot wun away fom berd. Them gonna chate her". In milder cases, there may be more subtle problems in understanding or using language, associated with difficulties in learning to read and spell. According to some authorities, 'developmental dyslexia' (specific reading retardation) is a milder version of SLI.

In the past few years, genetic factors have

been shown to be implicated in SLI (refs 5, 6). This has stimulated interest from linguists, who see the disorder as providing crucial evidence in support of Chomsky's views. If some children lack a capacity for learning language, and their deficiency has a genetic basis, this could indicate that specific genes are important for language acquisition, and these genes are not implicated in other cognitive abilities⁷. But such logic is dangerous, as can be illustrated by a simple analogy — nobody supposes that because a single genetic mutation can cause muscular dystrophy, there is a gene for 'walking'. In the same way, many researchers argue that, far from being a manifestation of damage to a specialized language-learning system, SLI may be the end result of impairment of some general perceptual or cognitive process that is particularly important for language learning.

So what kind of nonspecific deficit might lead to SLI? A strong candidate is impaired auditory perception. Deafness is always ruled out when diagnosing SLI, but that does not necessarily mean that the child's processing of auditory signals is normal. To learn language, a child must be able to detect when sounds occur (which is what a hearing test establishes), and also to discriminate between, and categorize, sounds. Children with SLI tend to do poorly in simple tasks that require them to discriminate between brief tones, or series of sounds occurring in rapid succession. For instance, a child may be trained to press one key on hearing a high tone and another on hearing a low tone. When presented with tone pairs, such as high-high or high-low and asked to press the corresponding sequence of keys, many children with SLI can perform perfectly with a 300-ms interval between tones. But their

performance drops to chance when a much shorter interval, for example 40 ms, is used. In contrast, normally developing children have little difficulty with such brief intervals⁸. Because many consonants are differentiated only in terms of a brief portion of the acoustic signal, which may be as short as 40 ms (see box), a difficulty in processing auditory signals that occur in rapid succession might plausibly account for difficulties in language learning.

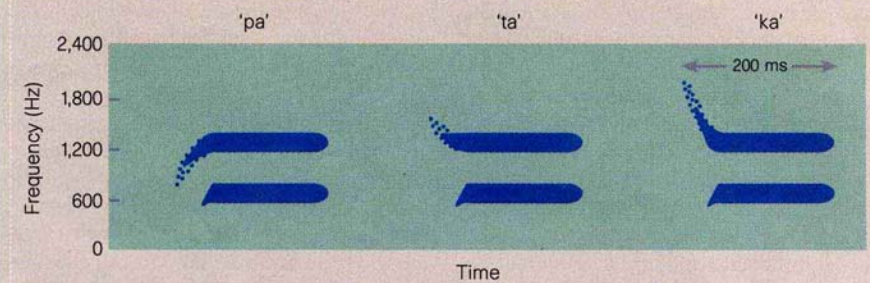
The 'auditory impairment' view of SLI has not gone unchallenged. Difficulties in discriminating non-verbal auditory stimuli in SLI are far less consistent and robust than problems with linguistic stimuli (such as distinguishing 'pa' and 'ka'). So, non-linguistic auditory impairments could be a correlated symptom of SLI, rather than a core deficit^{9,10}.

The study by Wright *et al.*³ now provides crucial evidence in this debate, because they show specific auditory abnormalities in children with SLI, using non-linguistic stimuli.

Studies on normal adults have established that detection of a brief tone can be affected by the presence of another auditory stimulus occurring immediately afterwards — a phenomenon known as backward masking. This has been explained in terms of a model in which a masking sound terminates perceptual processing of the target sound before enough information has accumulated for accurate detection: time is required to process the information in the first stimulus, and it cannot be properly perceived if another stimulus follows quickly¹¹. Wright *et al.* show that there is massively enhanced backward masking in children with SLI. Whereas control children could reliably detect a 45-decibel, 1,000-Hz tone that was immediately followed by a broad-band masking noise (a hissing sound), children with SLI could not: the volume of the stimulus tone had to be increased to nearly 90 decibels for them to achieve control performance levels. This could not be attributed to any general elevation of auditory thresholds or to attentional lapses, because the two groups of children performed at equivalent levels when, rather than preceding the noise, a tone was embedded in it. The authors also show that the extent to which a following noise interferes with detection of a tone depends on the overlap in frequency between the two.

Excessive backward masking could play an important role in causing SLI, by interfering with development of the ability to discriminate between speech sounds such as 'pa', 'ta' and 'ka', where a brief consonant portion is immediately followed by a vowel. Critics of the 'auditory impairment' theory will nevertheless argue that Wright *et al.* provide only correlational evidence, and that

Acoustic cues used in speech perception



The spectrograph is a device that converts speech signals into a three-dimensional visual representation, the spectrogram. Variations in intensity are represented by the darkness of the trace. Variations in frequency are represented on the y-axis, and time is

shown on the x-axis. In the 1950s, a 'pattern playback' was devised, which converted spectrograms back to an auditory form. This enabled researchers to modify the spectrogram, or even to create hand-painted spectrograms to investigate the auditory perceptual correlates

of particular changes in the signal. The spectrographic patterns shown (adapted from ref. 13) are used to synthesize the sounds for the syllables 'pa', 'ta' and 'ka', and they are identical except for a very brief initial portion lasting some 40 ms.

D.V.M.B.

they have not conclusively shown that auditory impairment is a cause of SLI. Instead, it is possible that auditory-masking deficits reflect a general immaturity or dysfunction of the nervous system. We know, for instance, that children with SLI perform poorly in other sensorimotor tasks that are less obviously related to their linguistic deficits, such as visual discrimination and motor co-ordination¹². So, enhanced auditory masking could be just another correlated deficit.

Another possibility is that excessive backward masking might be a consequence, rather than a cause, of poor speech perception: for example, by discriminating speech sounds, a child may practise perceiving rapid, successive auditory stimuli. Auditory masking might then be an index of expertise in speech perception. To rule this out, normal backward masking would need to be shown in people with limited language abilities who did not suffer from SLI. An obvious test would be to compare normally developing 3- or 4-year-olds with 7- to 8-year-olds whose language abilities are at a similar level because of SLI. The most convincing way to show that abnormal backward masking causes SLI would be to develop a training method for improving a child's resistance to auditory backward masking, and to show that this had a beneficial effect on language learning.

Although the study by Wright *et al.*³ does not prove that auditory masking is causally implicated in SLI, it increases the chances that the debate between the auditory and linguistic theories will soon be resolved. It shows that auditory impairment in SLI can

be studied in terms of a well-established psychophysical model, that of backward masking, which has potential for clinical application in diagnosing and, possibly, in treating children with SLI. So confining children in a barrel is probably not the only factor that can impair their language learning — enhanced backward masking may ensure that, even if they are exposed to adequate language, children fail to extract the critical cues for speech perception. □

D. V. M. Bishop is at the Medical Research Council Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF, UK.

1. Pinker, S. *Language Learnability and Language Development* (Harvard Univ. Press, Cambridge, MA, 1984).
2. Chomsky, N. *Rules and Representations* (Blackwell, Oxford, 1980).
3. Wright, B. A. *et al.* *Nature* **387**, 176–178 (1997).
4. Bishop, D. V. M. in *Child and Adolescent Psychiatry* (eds Rutter, M., Hersov, L. & Taylor, E.) 546–568 (Blackwell Scientific, Oxford, 1994).
5. Tomblin, J. B. & Buckwalter, P. R. in *Specific Language Impairments in Children* (eds Watkins, R. & Rice, M.) 17–34 (Brookes, Baltimore, 1994).
6. Bishop, D. V. M., North, T. & Donlan, C. *Dev. Med. Child Neurol.* **37**, 56–71 (1995).
7. Jackendoff, R. *Patterns in the Mind* (Harvester Wheatsheaf, Hemel Hempstead, 1993).
8. Tallal, P., Miller, S. & Fitch, R. H. in *Temporal Information Processing in the Nervous System* (eds Tallal, P., Galaburda, A. M., Llinas, R. R. & von Euler, C.) 27–47 (Ann. NY Acad. Sci., New York, 1993).
9. Studdert-Kennedy, M. & Mody, M. *Psychon. Bull. Rev.* **2**, 508–514 (1995).
10. Mody, M., Studdert-Kennedy, M. & Brady, S. J. *Exp. Child Psychol.* **64**, 199–231 (1997).
11. Massaro, D. W. & Burke, D. *Dev. Psychol.* **27**, 85–96 (1991).
12. Powell, R. & Bishop, D. V. M. *Dev. Med. Child Neurol.* **34**, 755–765 (1992).
13. Liberman, A. M., Ingemann, E., Lisker, L., Delattre, P. & Cooper, F. S. *J. Ac. Soc. Am.* **31**, 1490–1499 (1959).

Wandering why

Richard G. Gordon

T rue polar wander is the slow, secular shift of the entire solid Earth, relative to its rotation axis, in response to Earth's changing distribution of mass. In recent years the emphasis of dynamical models for true polar wander has turned from deglaciation, which is responsible for much of the wander this century, to the processes that drive polar wandering over millions of years^{1,2}. Now, Steinberger and O'Connell (page 169 of this issue³) have produced an actualistic model that makes predictions of true polar wander over the past 65 million years (Myr), and they agree strikingly well with the wander observed from palaeomagnetism and plate reconstructions relative to hotspots⁴.

Eight decades of observations from the International Latitude Service indicate, and two decades of space geodetic observations verify, that the solid Earth is moving relative to its rotation axis at a rate of about 10 cm yr^{-1} , or 1° Myr^{-1} (ref. 5). Estimates of true polar wander over millions of years are very uncertain and indicate a net drift of about 10° over Cenozoic time (the past 65 Myr), giving an average rate of drift of about $0.15^\circ \text{ Myr}^{-1}$, but with drift as fast as about $0.5^\circ \text{ Myr}^{-1}$ occurring in the past 10 Myr or so (ref. 6).

True polar wander is a consequence of the changing location of Earth's maximum principal axis of non-hydrostatic moment of inertia (that is, where the maximum principal axis would be on a non-rotating Earth). The diurnal rotation causes a bulge at the Equator and a flattening of the Poles resulting in the equatorial radius exceeding the polar radius by about 21 km. Without the diurnal rotation this difference would shrink dramatically, and the average equatorial radius would exceed the average polar radius by about only 100 m. Here the tail wags the dog — as the 100-m non-hydrostatic departures from sphericity migrate with time, the 21-km hydrostatic bulge gradually follows.

Over millions of years, surface features are unlikely to significantly perturb Earth's non-hydrostatic shape and moment of inertia. Nearly everywhere, the lateral mass anomalies implied by the topographic highs and lows of Earth's surface are cancelled out at depths of less than about 100 km, because the crust and upper mantle lack the strength to sustain appreciable shear stress on vertical planes. The lack of correlation between the geoid and nearly all surface topographic features (except deep-sea trenches) is direct evidence for this shallow compensation. Probably the most important source of the mass anomalies responsible for the departures of the non-hydrostatic figure from sphericity are the cold, dense, subducting slabs deliv-

ered into the mantle at deep-sea trenches². Their changing geometry indicates that the non-hydrostatic figure evolves with a characteristic time of about 50 Myr (ref. 7).

By how long does the 21-km bulge lag behind the 100-m bulge? Interpretations in the 1960s favoured the idea that Earth's mantle was viscous enough to cause a very long time lag. The tiny non-hydrostatic bulge was believed to have been inherited from tens or hundreds of millions of years before, when Earth's diurnal rotation was much faster than it is now (the Earth has gradually been decelerating over its history because of tidal frictional dissipation of its rotational energy). In 1969, Goldreich and Toomre⁸ showed that the tiny departures from sphericity of the non-hydrostatic figure of the Earth are as expected if the hydrostatic bulge tracks them in a geologically negligible amount of time. Work in the early 1990s revived the hypothesis of a slow response of the bulge¹, but the most recent results — those of Richards *et al.*², published earlier this year, and of Steinberger and O'Connell³ — show that the hydrostatic bulge adjusts quickly.

Steinberger and O'Connell show, for example, that the rotation axis would lag the inertia axis by less than 1° assuming a reasonable upper limit to the viscosity of the lower mantle. Simulations indicate that rapid and large-amplitude polar wander, with occasional swings of about 90° , is to be expected on an Earth-like planet⁸. But the palaeomagnetic and plate-reconstruction record indicate that the swings over the past 100 Myr are much smaller than this, no more than 10° . The absence of large swings might imply that the viscosity of the lower mantle is high enough to slow true polar wander, but both Richards *et al.* and Steinberger and O'Connell find instead that the migrations of the axis of maximum non-hydrostatic moment of inertia have been small. Richards *et al.* explain the smallness of true polar wander over the past 100 Myr by the slow rate of change of the large-scale pattern of plate-tectonic motions and associated geometry of subducting slabs during the past 200 Myr.

So how do the two studies differ? Whereas Richards *et al.*² calculate the changing non-hydrostatic moment of inertia due to the evolving geometry of subducting slabs of oceanic lithosphere, Steinberger and O'Connell³ take a complementary approach. They start with the present distribution of density anomalies estimated from anomalies in the velocity of seismic-wave propagation and an empirical relation between density and velocity. They trace these anomalies backwards in time in a modelled flow that is driven by buoyancy anomalies and is consistent



100 YEARS AGO

English weather is as fruitful a subject for composition as it is a theme for conversation. Like many other people, Mr. C. A. Whitmore, M. P., is a devoted student of our meteorology, so that what he writes about in the May number of the *National Review* is worth reading. Weather fallacies have been exposed times without number, but they are so deeply rooted in the minds of the unscientific that it may be doubted whether they will ever be completely eradicated. Mr. Whitmore throws doubt upon the popular impression that the changes of the moon synchronise with marked changes of weather. The few facts he states as to weather and lunar phases since the beginning of summer, ought to convince people that their faith in the influence of the moon is misplaced. Another very common idea is that a heavy dew at night presages a fine day on the morrow, whereas it only indicates that the sky is clear and conditions are favourable for the deposition of dew. At certain times of the year a heavy dew is a sign of unstable rather than of stable weather. A luxuriant crop of berries in the autumn is said to forebode a severe winter; but the people who believe this, forget, or do not know, that the berries tell of conditions which have passed rather than of those to come.

From *Nature* 6 May 1897.

50 YEARS AGO

There is a large gap in our knowledge which must be bridged before we can take decisive steps from present aircraft subsonic speeds to supersonic speeds. ... Normally one would expect to get a large part of this information from wind tunnels, and although the capabilities of existing wind tunnels will be used to the limit, nobody knows how to design or to use a practicable wind tunnel at or about the speed of sound. ... This is the main reason for going into the air, and a series of experiments has been planned to use flying model aircraft launched from a great height from another aircraft and powered by rockets.

From *Nature* 10 May 1947.

Many more abstracts like these can be found in *A Bedside Nature: Genius and Eccentricity in Science, 1869–1953*, a 266-page book edited by Walter Gratzer. Contact David Plant (e-mail: subscriptions@nature.com).

with the sources and sinks imposed by the motion between the surface tectonic plates.

The two approaches give predicted paths of true polar wander that are broadly similar, especially in giving similar 10° amplitudes of true polar wander over the past 65 Myr. But they differ in direction, with the path calculated by Steinberger and O'Connell³ being in excellent agreement with the observed path of true polar wander and the path of Richards *et al.*² being in only fair agreement. Whereas the latter reflects the changing density distribution due to the evolving geometry of subducting slabs, the former implicitly captures all sources of the changing density distribution. The difference in the two calculated paths could possibly be due to inaccuracies in the modelled history of subduction, but it may be telling us something important about other smaller yet significant sources of evolving mantle density structure. Possibilities to be considered by geodynamicists include upwelling mantle plumes, and downwelling cold mantle at locations other than subduction zones (for example by convective destabilization of mantle lithosphere beneath horizontally shortened continental plateaux).

The plate reconstructionists also have work to do. Models of the history of subduction can be no better than the plate reconstructions on which they are based, and debates about many features of circum-Pacific plate motion are still far from settled. (Unfortunately, the nature of some pre-Cenozoic features may never be known because of subduction of the sea floor that records the older plate motions.) Moreover, published uncertainties on paths of true polar wander include the uncertainties in the palaeomagnetic data but neglect those, possibly larger, uncertainties in the reconstructions of the plates relative to the hotspots. So it is necessary to incorporate these uncertainties to ascertain whether the differences between the observed path of true polar wander and that calculated by Richards *et al.*² are significant before giving too much weight to the difference.

Palaeomagnetism has contributions to make as well. First, better palaeomagnetic data for the Pacific plate and the surrounding continents is likely to be the ultimate arbiter between the distinctly different plate reconstructions that have been proposed for the Pacific Ocean basin. Moreover, estimated paths of true polar wander are based entirely on data from the continents and, to be really convincing, also require data from the Pacific. Second, the observed path of true polar wander is a highly smoothed version of a possibly undersampled signal. Palaeomagnetic data of higher accuracy and age resolution would place a more stringent bound on viscosity of the lower mantle than is now possible.

Palaeomagnetically observed true polar wander is similar in direction to true polar wander observed this century. This agreement would be expected if the long-term wander was much faster than that caused by deglaciation.

Such is not the case, however, and the agreement is mainly a coincidence⁹. Although the $0.6^\circ \text{ Myr}^{-1}$ along 40° W of true polar wander over the past 10 Myr is much faster than the average rate over many tens of millions of years⁵, it is still slower than true polar wander this century, which is about $1.1^\circ \text{ Myr}^{-1}$ along 70° W (ref. 5). Steinberger and O'Connell's model predicts an even slower rate for the long-term wander, about $0.4^\circ \text{ Myr}^{-1}$ along 24° W over the past 1 Myr. This predicted rate, if accurate, can be subtracted from the historically observed wander to obtain an estimate of the wander due to deglaciation of $0.9^\circ \text{ Myr}^{-1}$ towards 90° W , which is still faster, perhaps much faster, than the long-term rate of wander. □

Richard G. Gordon is in the Department of Geology and Geophysics, Rice University, Houston, Texas 77005, USA (e-mail: rgg@geophysics.rice.edu).

1. Spada, G., Ricard, Y. & Sabadini, R. *Nature* **360**, 452–454 (1992).
2. Richards, M. A., Ricard, Y., Lithgow-Bertelloni, C., Spada, G. & Sabadini, R. *Science* **275**, 372–375 (1997).
3. Steinberger, B. & O'Connell, R. J. *Nature* **387**, 169–172 (1997).
4. Morgan, W. J. in *Oceanic Lithosphere* (The Sea Vol. 7) (ed. Emiliani, C.) 443–487 (Wiley, New York, 1981).
5. Gross, R. S. in *IERS Tech. Note 17* (ed. Charlot, P.) C5–C12 (Observatoire de Paris, Paris, 1994).
6. Besse, J. & Courtillot, V. *J. Geophys. Res.* **96**, 4029–4050 (1991).
7. Ricard, Y., Richards, M., Lithgow-Bertelloni, C. & Le Stunff, Y. *J. Geophys. Res.* **98**, 21895–21909 (1993).
8. Goldreich, P. & Toomre, A. *J. Geophys. Res.* **74**, 2555–2567 (1969).
9. Gordon, R. G. *J. Geophys. Res.* **100**, 24367–24392 (1995).

Protein transport

A fusion of new ideas

Jason B. Bock and Richard H. Scheller

Eukaryotic cells rely on a series of membrane-bound organelles to compartmentalize biochemical reactions and to regulate the secretion and localization of proteins. Although each organelle has a unique molecular composition, giving it a

distinct functional identity, membranes and proteins are continuously shuttled between compartments. This process is mediated by transport vesicles that bud from the membranes of donor compartments and then fuse with acceptor membranes (Fig. 1).

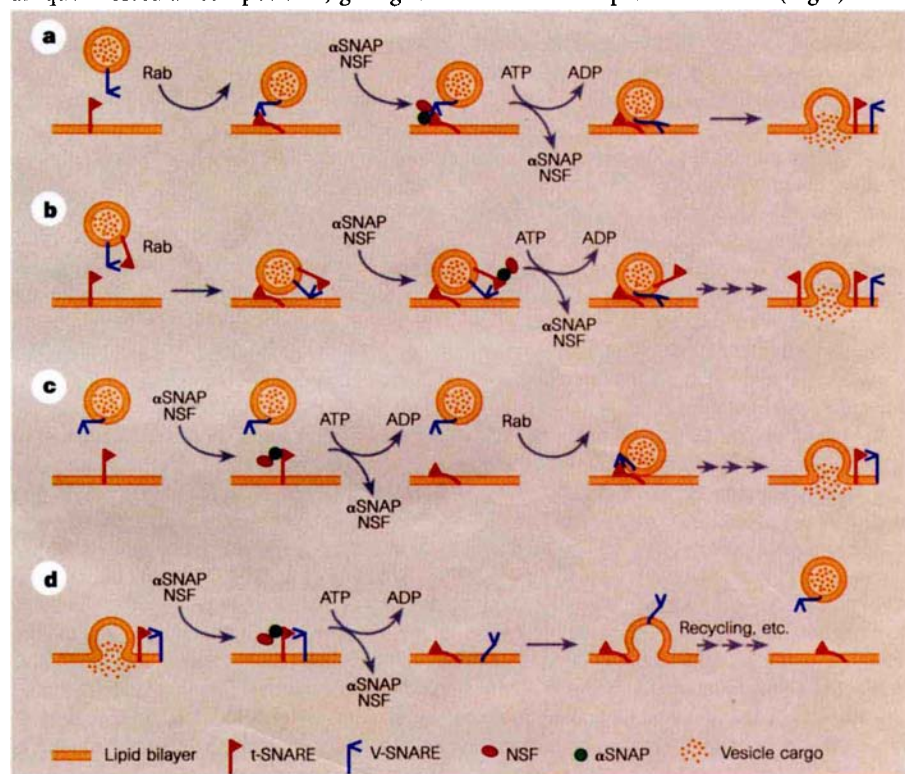


Figure 1 a, The low-molecular-weight Rab-family GTPases are involved in vesicle target recognition, followed by the action of α -SNAP and NSF on a v- and t-SNARE complex. NSF-mediated SNARE rearrangement leads to membrane fusion, and Nichols *et al.*¹ have shown that v- and t-SNAREs are required on opposing membranes for vacuole fusion in yeast. b, The v- and t-SNAREs were originally thought to be localized to vesicle and target membranes, respectively, but a fraction of the t-SNAREs are also found on vesicles. This allows potentially physiologically important v- and t-SNARE pairing to occur on the same membrane. c, α -SNAP and NSF act on the t-SNARE alone, altering its conformation to allow Rab-mediated vesicle targeting. A series of undefined steps leads to membrane fusion. d, After membrane fusion, α -SNAP and NSF act to untangle the SNARE proteins, which are then segregated to vesicle and acceptor membrane compartments for another round of docking.

How do transport vesicles recognize the appropriate acceptor membrane? Recognition is thought to occur, at least in part, by the specific binding of proteins on the vesicle (v-SNAREs) to distinct partners on target membranes (t-SNAREs). This has been termed the SNARE hypothesis and, on page 199 of this issue, Nichols *et al.*¹ examine some of the key tenets of the model.

The SNARE hypothesis predicts that pairing of v- and t-SNAREs is regulated by low-molecular-weight proteins known as Rabs, which hydrolyse GTP. After the v- and t-SNAREs have paired, the soluble ATPase, *N*-ethylmaleimide-sensitive factor (NSF), binds the SNARE complex through the soluble NSF-attachment protein (α -SNAP) and hydrolyses ATP. This results in reorganization of the complex and membrane fusion^{2,3} (Fig. 1a).

Using an elegant combination of genetics and biochemistry, Nichols *et al.* have manipulated the SNARE composition of the donor and acceptor membranes. Consistent with earlier hypotheses, they show that a prerequisite for fusion of two membranes is a t-SNARE on one membrane and a v-SNARE on the other. But, in conflict with the SNARE hypothesis, their experiments reveal a role for NSF prior to docking — before the acceptor and donor membranes ever see each other^{1,4}.

The seed of the SNARE hypothesis was planted in the late 1980s and early 1990s, with a molecular dissection of the proteins involved in synaptic vesicle transport in neurons (reviewed in ref. 5). Studies of neurotransmitter-filled vesicles from the brain uncovered a prototype v-SNARE, known as VAMP (ref. 6) or synaptobrevin. The prototype t-SNARE, called syntaxin 1, was characterized as a nerve-terminal protein that associates with proteins on the synaptic vesicle. This interaction between proteins on opposing membranes was proposed to mediate docking of the vesicle at the plasma membrane⁷.

It soon became clear that such pairing between v- and t-SNAREs defined a model that encompassed many different vesicle-trafficking steps throughout the secretory pathway, in cells as evolutionarily distant as yeast and neurons⁸. Studies in yeast identified a set of v- and t-SNAREs that mediate transport between the endoplasmic reticulum and the Golgi apparatus^{9,10}. A different set of SNAREs was found to underlie shuttling between the Golgi apparatus and vacuole (lysosome)¹¹. The emerging concept was that specific pairs of vesicle and target membrane proteins — in general, of the VAMP and syntaxin families, respectively — mediate the fidelity of vesicle trafficking. Moreover, the cytosolic proteins α -SNAP and NSF, which promiscuously interact with SNARE pairs, were thought to mediate this membrane fusion throughout the secretory pathway³ (Fig. 1a).

No sooner was this hypothesis put forth, than potentially conflicting data were generated. First, although vesicle fusion occurs at

Mammalian SNAREs

Protein	Accession number	Protein	Accession number
Syntaxin 1a	M95734	VAMP 1	M24104
Syntaxin 1b	M95735	VAMP 2	M24105
Syntaxin 2	L20823	VAMP 3	S63830 (Cellubrevin)
Syntaxin 3	L20820	VAMP 4	D86817 AA197391
Syntaxin 4	L20821	VAMP 5	AA222692
Syntaxin 5	L20822	VAMP 6	W69164
Syntaxin 6	U56815	VAMP 7	X96737
Syntaxin 7	D60600 AA081523	VAMP 8	AA049140
Syntaxin 8	AA111025 W41301	Rsec22a	U42209
Syntaxin 9	AA150357	Msec22b	U91538
Syntaxin 10	N35629 W24393	Hsec22c	H59647
Syntaxin 11	AA227632	Rbet1a	U42755
Syntaxin 12	R29508	Mbet1b	W75334 W83047
Syntaxin 13	AA167677	Membrn	U91539
Syntaxin 14	T08774	GOS-28	U49099
Syntaxin 15	AA244750	Hykt6a	H18270 H23796
Syntaxin 16	AA100145	Hykt6b	AA016381
SNAP-25	M22012		
SNAP-23	U55936		

Differential expression patterns, subcellular localization and protein-protein interactions between v- and t-SNAREs may determine the organization of membrane compartments in cells, by controlling the specificity of vesicle trafficking between these compartments. If SNAREs work in this way, then there must be more of them than have been

identified to account for the known anterograde and retrograde trafficking steps. As a first step towards understanding the complete complement of SNAREs in mammalian species, we searched the NCBI EST database with the sequences of known SNAREs. Relevant sequences were determined through alignment of prospective, with known, SNAREs using the BESTFIT and

PILEUP programs. Prospective sequences were randomized and again aligned with known SNAREs. In all cases the original quality score was at least ten standard deviations higher than that obtained after randomization. Bold type indicates newly identified SNAREs. All sequences can be found at NCBI Entrez Web site: <http://www3.ncbi.nlm.nih.gov/Entrez/> **J.B.B. & R.H.S.**

specialized regions of the plasma membrane, the t-SNAREs are not localized to specific regions — they are distributed over most of the cell membrane. Perhaps even more troubling, a fraction of the synaptic t-SNAREs were found on vesicles, and these were suggested to be the physiologically active component (Fig. 1b). Second, neurotoxin cleavage of SNARE proteins, which renders them nonfunctional, does not deplete the pool of vesicles that seem to be docked and ready for membrane fusion. Finally, physiological studies indicated that the last requirement for ATP hydrolysis occurred considerably before the fusion of vesicles, making it difficult to understand how NSF could directly drive the membrane-fusion event.

To address these discrepancies, Nichols *et al.*¹ used a biochemical *in vitro* assay¹². Vacuolar membranes are isolated from two different yeast strains, and the rate of fusion is monitored through a biochemical reaction that only occurs when the internal vacuolar contents of the two strains are intermixed. Vacuole fusion is considered to be 'homo-

typic', because the donor and acceptor membranes have identical protein constituents, including their v- and t-SNAREs. The power of the assay lies in the fact that the constituents of the vacuolar membranes of each yeast strain can be altered independently.

The authors identified the v- and t-SNARE proteins that are required for homotypic vacuole fusion by searching the recently completed yeast genome sequence for VAMP- and syntaxin-related sequences. They then created strains that were deficient in either of these newly found vacuolar v- and t-SNAREs. When vacuoles containing either v-SNAREs or t-SNAREs alone were mixed, fusion was reduced to near background levels. Fusion was also greatly reduced when one set of vacuoles contained both v- and t-SNAREs, and the other set lacked both proteins. But when one set of vacuoles contained only the v-SNARE, and the other set contained only the t-SNARE, efficient fusion occurred. So, consistent with the SNARE hypothesis, Nichols *et al.* concluded that v- and t-SNAREs are required, on opposing membranes, for

vacuole fusion. Further studies are now needed to confirm that the v- and t-SNAREs that they used do, in fact, specifically interact with each other, perhaps through direct binding (as has been shown for v- and t-SNARE pairs in other trafficking steps).

The authors did not stop there; however, the next set of results did not fit the SNARE hypothesis. But, they may help to explain some of the earlier contradictory data, which indicated that the action of NSF, mediated by ATP hydrolysis, occurs early in the biochemical pathway leading to fusion⁴ and is followed by the action of Ypt7p, a member of the Rab family of small GTPases¹³. Nichols *et al.* found that α -SNAP and NSF were only required before the vacuoles were mixed, even in the artificially created heterotypic case, where individual fusion partners contained only v- or only t-SNAREs. So, they believe that α -SNAP and NSF are needed to prime individual SNAREs before docking, probably to change their conformation to an activated state (Fig. 1c).

If the order of events occurs as depicted in Fig. 1c, how is the very stable SNARE complex broken apart for another round of vesicle docking and fusion? Perhaps NSF catalyses this event as well (Fig. 1d). Other issues, including the membrane-fusion process itself, remain poorly understood at the biochemical level. Furthermore, many other proteins are critical for vesicle trafficking, yet their functions are still largely unknown. In the brain, regulation of synaptic vesicle trafficking through SNARE proteins is likely to be critical in modulating neurotransmitter release — a process that is implicated in the formation of memories. Perhaps, then, the ultimate challenge will be to understand the biochemical mechanisms through which the neural activities that are generated through our experiences change SNARE-mediated synaptic properties, to contribute to the storage of memories. □

Jason B. Bock and Richard H. Scheller are at the Howard Hughes Medical Institute, Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, California 94305, USA.

- Nichols, B. J., Ungermann, C., Pelham, H. R. B., Wickner, W. T. & Haas, A. *Nature* **386**, 199–202 (1997).
- Söllner, T. *et al.* *Nature* **362**, 318–324 (1993).
- Söllner, T., Bennett, M. K., Whiteheart, S. W., Scheller, R. H. & Rothman, J. E. *Cell* **75**, 409–418 (1993).
- Mayer, A., Wickner, W. & Haas, A. *Cell* **85**, 83–94 (1996).
- Sudhof, T. C. *Nature* **375**, 645–653 (1995).
- Trimble, W. S., Cowan, D. & Scheller, R. H. *Proc. Natl Acad. Sci. USA* **85**, 4538–4542 (1988).
- Bennett, M. K., Calakos, N. & Scheller, R. H. *Science* **257**, 255–259 (1992).
- Bennett, M. K. & Scheller, R. H. *Proc. Natl Acad. Sci. USA* **90**, 2559–2563 (1993).
- Lian, J. P. & Ferro-Novick, S. *Cell* **73**, 735–745 (1993).
- Hardwick, K. & Pelham, H. R. B. *J. Cell Biol.* **119**, 513–521 (1992).
- Becherer, K. A., Rieder, S. E., Emr, S. D. & Jones, E. W. *Mol. Biol. Cell* **7**, 579–594 (1996).
- Haas, A., Conradt, B. & Wickner, W. *J. Cell Biol.* **126**, 87–97 (1994).
- Mayer, A. & Wickner, W. *J. Cell Biol.* **136**, 307–317 (1997).

Evolutionary genetics

The muddle about mutations

Joel R. Peck and Adam Eyre-Walker

Deleterious mutations are genetic changes that decrease fertility or the probability of survival — that is, they decrease fitness. Most evolutionary biologists believe that deleterious mutations that decrease fitness by more than one per cent are common in multicellular plants and animals, arising at a rate of about one new mutation for every offspring. This belief is mainly based on experiments using the fruitfly *Drosophila melanogaster*, which were carried out by Terumi Mukai and his colleagues¹. If their estimates are both correct and general, then deleterious mutations can have a significant evolutionary impact. For example, they may strongly affect the evolution of sexual reproduction, mate choice and neutral molecular variation^{2,3}. But three recent reports^{4–6}, culminating in a paper by Peter Keightley and Armando Caballero⁶ in the 15 April issue of *Proceedings of the National Academy of Sciences*, have shaken the foundations of the experimental edifice on which current estimates of the rate (and magnitude) of deleterious mutations rest.

The first of the new studies was by Jesús Fernández and Carlos López-Fanjul⁴. Like Mukai *et al.* they used *Drosophila*, but their data seem to indicate that deleterious mutations cause much less damage than would be expected from Mukai *et al.*'s estimates. This might be because deleterious mutations are rare, or because mutations usually decrease fitness by much less than one per cent. In another study, Peter Keightley⁵ raised serious questions about the basic methods that were used in Mukai *et al.*'s experiments. Now, in collaboration with Armando Caballero⁶, Keightley has completed a study of mutation in the nematode worm *Caenorhabditis elegans*. The results are roughly in accord with those of Fernández and López-Fanjul in that they show surprisingly little damage as a result of deleterious mutations.

It is relatively easy to measure the rate of recessive lethal mutations. These are mutations that cause death when they are homozygous — that is, when offspring get the mutation from both parents. It is much more difficult to measure the rate and effect of mild deleterious mutations, which often do not kill, even in the homozygous form (although they do make survival and/or reproduction less likely). The current debate centres around these mild deleterious mutations, which are thought to be much more common than the recessive lethal sort. Keightley and Caballero⁶ suggest that the damage caused by mild deleterious muta-

tions in *C. elegans* is about 100 times less than calculated by Mukai *et al.* for *Drosophila*.

How can we explain these discrepancies? The most obvious possibility is that mutation rates vary widely between different populations of multicellular plants and animals. Perhaps experiments that have shown high rates of mutation have used populations that are subject to extraordinarily high mutation rates because of the action of mobile genetic elements. Or perhaps *C. elegans* has a particularly low rate of mutation. But *C. elegans* and *Drosophila* have similar numbers of genes, and the per-gene mutation rates also seem to be similar. Moreover, the rate of recessive lethal mutations in *C. elegans* seems to be only about three times lower than that in *Drosophila*⁶.

If the rate of lethal recessive mutations is similar in *C. elegans* and *Drosophila*, why do mild deleterious mutations seem to occur much less frequently — or have much smaller effects — in *C. elegans*? It is possible that Mukai *et al.*'s estimates are flawed, and that the rate (or effect) of deleterious mutations in most multicellular plants and animals is much smaller than is generally believed. Mukai *et al.*'s experimental procedures may be the source of the problem. Mild deleterious mutations are difficult to detect in isolation, so Mukai *et al.*'s method allowed them to accumulate on a particular chromosome by protecting them from the action of natural selection. This was done by keeping the mutations in heterozygous form (that is, by ensuring that they were inherited from one parent, but not from the other). In addition, the size of each experimental population was kept very small, to decrease the effectiveness of natural selection, and the flies were kept under good conditions, except when they were being tested for the effects of mutations. Then, during the tests, the fitness-altering effects of the chromosome that had been kept protected were compared with the fitness-altering effects of a control chromosome.

Keightley⁵ believes that, as a result of natural selection, the control chromosome may have improved over the course of the experiment, causing an apparent (but not real) decline in the quality of the chromosome that had been protected from selection. One difficulty with this interpretation is that Mukai *et al.*'s estimates in *Drosophila* do not depend solely on comparisons with a control chromosome. They are supported by the 'order method', which derives its estimates using variation between different repetitions of the experiment¹, and these estimates

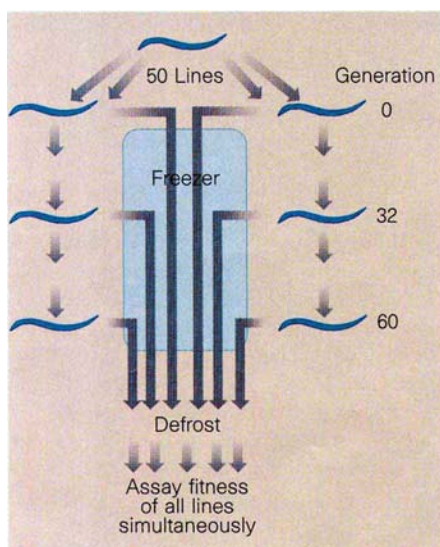


Figure 1 Keightley and Caballero⁶ used the self-fertilizing nematode *C. elegans* to measure the effect of mild deleterious mutations. They established 50 lines from the progeny of a single individual, and these were propagated for 60 generations. In each generation, a single, randomly selected larva was used to propagate each line. This breeding scheme minimizes the effect of natural selection. A sample of individuals was preserved alive by freezing at generations 0, 32 and 60, and frozen individuals were revived at the end of the experiment for simultaneous fitness assays.

roughly agree with those obtained using a control chromosome (although Keightley has questioned the validity of this finding⁵). In addition, plants that are highly inbred in their natural populations can be used to produce indirect estimates of the rate of mild deleterious mutations. Calculations from these studies^{7,8} suggest mutation rates that are in line with Mukai *et al.*'s estimates for *Drosophila*. However, the plant studies do not provide estimates of the magnitude of the effects of deleterious mutations.

A final possibility is that the new experimental results^{4,6} are misleading. Keightley and Caballero's experiment⁶, like those of Mukai *et al.*, depends on sheltering mild deleterious mutations from the effects of natural selection. But, because *C. elegans* is a self-fertilizing hermaphrodite, deleterious mutations cannot be protected from selection by keeping them in heterozygous form. So the authors had to rely mainly on the sheltering effects of small population sizes and good rearing conditions, as did Fernández and López-Fanjul⁴, who used inbred lines in their experiments on *Drosophila*. This should not matter much unless the damage caused by most deleterious mutations is considerably more than twice as large when they are homozygous, as when they are heterozygous. Furthermore, the population sizes used should have been small enough to allow many of the deleteri-

ous mutations to accumulate, even if they decreased fitness by as much as 40 per cent when in homozygous form^{6,9}. More troubling is that, during the tests of fitness, Mukai *et al.*'s flies were probably faced with more intense competition for food than were the animals in the new experiments. Moreover, the competing genotypes were more diverse than in the new studies. The conditions used in Mukai *et al.*'s tests may be more natural, and more likely to reveal differences in fitness. Keightley and Caballero address the concerns about competitive versus non-competitive tests, but the matter is not settled.

Previous estimates of the rate and magnitude of deleterious mutations are now seriously in doubt. Nevertheless, a solid understanding of these phenomena is within reach. It is not impossible to address any of the problems that have been discussed here, and studies could be extended to cover a wider range of populations and species. The procedure developed by Keightley and Caballero seems to be particularly promising for future investigations as it relies on freezing animals and then reviving them (Fig. 1), allowing the problems that are inherent in measuring experimental populations at a different time from the controls to be eliminated. (Note that the procedure of Fernández and López-Fanjul is susceptible to exactly this sort of problem.) With Keightley and Caballero's procedure, fitness can be assessed simultaneously in a population in which mutations have accumulated, and in the ancestors of that population (which act as a control). A similar advantage might be obtained by using organisms with very long-lasting eggs or seeds.

A better understanding of the characteristics of deleterious mutations is not only possible, it is imperative. Apart from the importance of the issues from a scientific perspective, the rate and magnitude of deleterious mutations have substantial implications for conservation genetics and, therefore, for preserving the diversity of life on Earth^{10,11}.

Joel R. Peck and Adam Eyre-Walker are at the Centre for the Study of Evolution, University of Sussex, Brighton BN1 9QG, UK.

1. Mukai, T. S., Chigusa, S. J., Mettler, L. E. & Crow, J. F. *Genetics* **72**, 339–355 (1972).
2. Charlesworth, B., Morgan, M. T. & Charlesworth, D. *Genetics* **134**, 1289–1303 (1993).
3. Kondrashov, A. S. *Nature* **336**, 435–440 (1988).
4. Fernández, J. & López-Fanjul, C. *Genetics* **143**, 829–837 (1996).
5. Keightley, P. D. *Genetics* **144**, 1993–1999 (1996).
6. Keightley, P. D. & Caballero, A. *Proc. Natl Acad. Sci. USA* **94**, 3823–3827 (1997).
7. Johnston, M. O. & Schoen, D. J. *Science* **267**, 226–229 (1995).
8. Charlesworth, D., Lyons, E. E. & Litchfield, L. B. *Proc. R. Soc. Lond. B* **258**, 209–214 (1994).
9. Crow, J. F. & Kimura, M. *An Introduction to Population Genetics Theory* (Burgess, Minneapolis, 1970).
10. Lande, R. *Conserv. Biol.* **9**, 782–791 (1995).
11. Lynch, M., Connery, J. & Burger, R. *Am. Nat.* **146**, 489–518 (1995).

Daedalus

Tall ships

The fastest-reacting control systems are inherently unstable. This is why a bicycle is so manoeuvrable, and why modern fighter aircraft are so unstable that they must be flown by computer. Daedalus is now carrying this message into new waters. He is designing an unstable boat.

Imagine, he says, a twin-hulled catamaran, turned through 90° about its axis of travel. One hull would then be under the water, the other would be held clear of it, and the two would be joined by vertical spars passing through the water surface. If the submerged hull had sufficient buoyancy to support the exposed one, the craft, though unstable, would be in formal hydrostatic balance.

This novel 'vertimaran' has many virtues. Most ships waste a vast amount of power making waves. But with almost no waterline, the vertimaran will make no waves. Properly streamlined, its lower hull will have relatively little viscous drag; its exposed upper hull will only feel wind resistance; so its fuel costs will be very low. Even better, the vertimaran will be wonderfully unaffected by rough seas. The cross-section of its waterline being negligible, passing waves will produce almost no change of upthrust. A wave crest would have to reach the upper hull, or its trough expose the lower one, before the vessel felt any perturbation.

But how to overcome the craft's instability? Left to itself, the vertimaran would simply fall over sideways, bringing both hulls back to the surface. Daedalus notes that many passenger vessels already have gyro-controlled stabilizer vanes under the water, twisting on command to oppose rolling forces. He is extending and improving this technique. The vertimaran's submerged hull will have several sets of active, servo-powered vanes, always keeping the lower hull vertically under the upper one, thus holding the craft upright.

Like a bicycle, the vertimaran will lose balance if it slows down. When it nears a port, ballast tanks in the submerged hull will be filled with water. The craft will sink down until the upper hull is afloat, converting it to a stable vessel. On leaving harbour with its new cargo, the tanks will be pumped out again; it will rise out of the water and speed on its way.

The vertimaran, with its high speed and utter stability, will be a perfect passenger vessel. Ferry companies and cruise lines will rush to adopt it. And a small one-seater version would make an ideal sporting craft, a sort of sea-going motorbike.

David Jones

Coral polyp expulsion

In coral reef environments that suffer regular physical disturbances, the fragmentation of established colonies is believed to be important in the recruitment (recolonization) of the local area by branching and free-living corals¹⁻⁴. High rates of wave action and sedimentation, for example, would favour fragmentation over the settlement and metamorphosis of sexually produced larvae¹. Nonetheless, massive and encrusting coral species, which are abun-

dant in sandy shallow tropical waters, show only minimal fragmentation¹. Here we report a mode of asexual reproduction in massive and encrusting corals, which enables them to sustain local populations in physically disturbed environments.

We have observed an unusual form of budding in massive and encrusting corals from shallow waters, particularly in *Favia favus* from the northern Red Sea and in *Oculina patagonica* from the Mediterranean coast of Israel. Individual polyps, including their calices, lift on elongated calcareous stalks (Fig. 1a) before detaching and settling elsewhere. The area of polyp expulsion (Fig. 2) regenerates within two weeks. The new propagule settles, its tissues spread over the stalk and onto the substrate (Fig. 1b), and the polyp forms a new colony through budding. In the laboratory, newly settled propagules of *O. patagonica* and *F. favus* developed and grew into new colonies within two months.

We also observed polyp expulsion *in situ*,

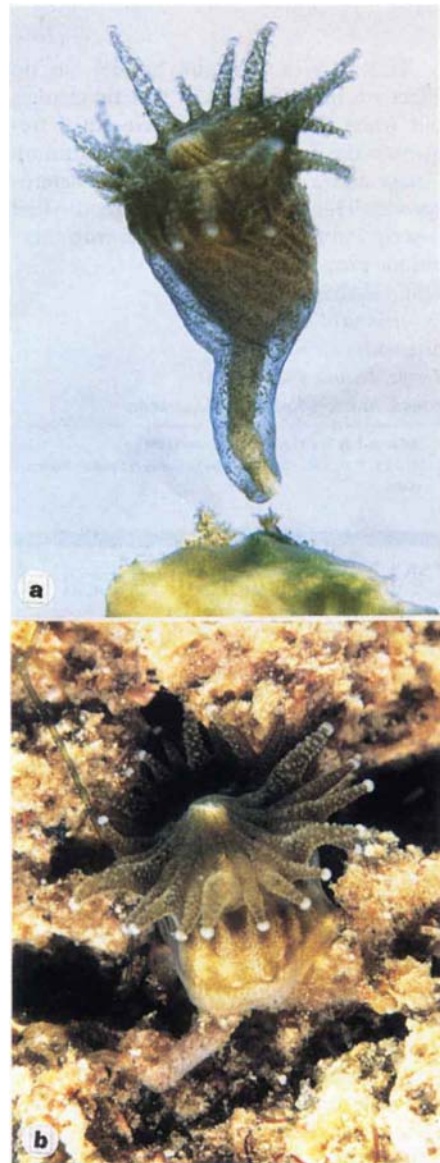


Figure 1 Polyp expulsion in the encrusting coral *Oculina patagonica*. **a**, An expelled polyp at the time of release. The tissue covering the calcareous stalk is visible ($\times 6.5$). **b**, A newly settled expelled polyp, with tissues beginning to spread over the stalk and eventually over the substrate ($\times 6.0$). The prevalence of expulsion was recorded along 10-m line transects⁸. Expulsion was observed in 34 of the 178 colonies of shallow-water *O. patagonica* recorded in 12 transects. All photographs by A. Shoob.

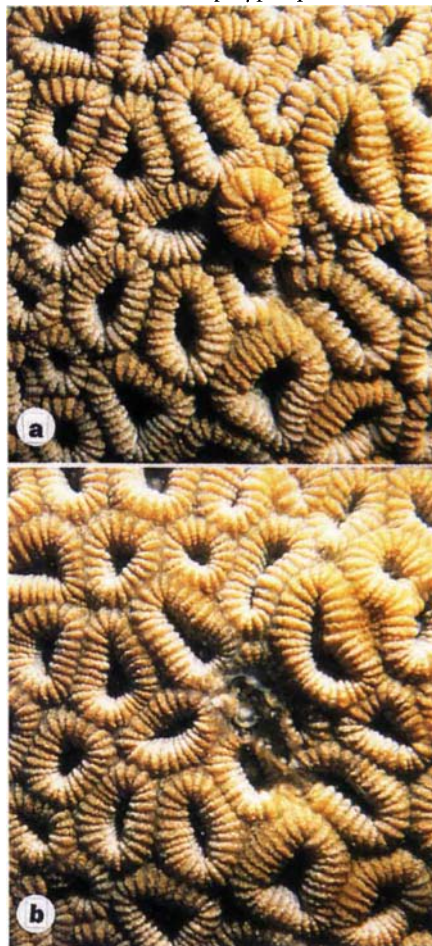


Figure 2 The massive coral *Favia favus*. **a**, An elevated polyp before expulsion; and **b**, the area of detachment immediately after expulsion ($\times 0.6$). Expulsion was seen in 10 of the 25 *F. favus* colonies recorded in nine transects.

in areas that suffered chronic levels of sedimentation, during the summer and autumn months when the water temperatures are highest (26 °C in Eilat and 29 °C in the Mediterranean). We did not see polyp expulsion in specimens from waters deeper than 4 m in the Red Sea and 7 m in the Mediterranean, indicating that this phenomenon might be unique to shallow water, where physical disturbances are more pronounced.

Histological sections revealed that during the peak reproductive season, polyps that were about to be expelled from the 'mother' colony contained no gonads, whereas surrounding polyps did. It is thus possible that the expelled polyp concentrates its stem-cell pool and energetic resources for use in reattachment and growth rather than for reproduction. Polyp expulsion may increase the survival rate of the parental genotype in a disturbed environment by ensuring that genotypically identical offspring are present at different stages in their life cycle⁵.

This phenomenon differs from the polyp 'bail-out' process, found in several branching pocilloporid coral species^{6,7}. In polyp bail-out, disintegrating corals release whole polyps from their calices, some with planulae, and the parent colony dies. In contrast, polyp expulsion occurs in physiologically healthy corals, and whole polyps, including their calices, are expelled. Unlike fragmentation, this process seems to be regulated by the colony and therefore may be considered a form of asexual reproduction.

As the main mode of reproduction in stony corals is sexual², the question arises as to how successful recruitment¹ occurs in chronically disturbed environments, where gametes and planulae may be swept away, dislodged or smothered by sediment. Our findings show that massive and encrusting stony corals which are abundant in such environments may overcome the difficulties of planular recruitment by reproducing asexually by polyp expulsion. As expelled polyps are larger than sexually derived propagules, they begin their sessile lives at the juvenile stage, avoiding the rigours of larval development, settlement and metamorphosis. This may be a form of size refuge, increasing the corals' chance of survival in disturbed environments.

E. Kramarsky-Winter, M. Fine, Y. Loya

Department of Zoology,

Tel Aviv University, Tel Aviv 69978, Israel

e-mail: yosiloya@ccsg.tau.ac.il

1. Highsmith, R. *Mar. Ecol. Prog. Ser.* 7, 207-226 (1982).
2. Harrison, P. L. & Wallace, C. C. in *Ecosystems of the World 25: Coral Reefs* (ed. Dubinski, Z.) 133-206 (Elsevier, Amsterdam, 1990).
3. Scheer, G. *Naturwiss. Ver. Darmstadt Ber.* 1958/59, 50-60 (1959).
4. Rosen, B. R. & Taylor, J. D. *Science* 166, 119-121 (1969).
5. Orive, M. E. *Am. Nat.* 145, 90-108 (1995).
6. Sammarco, P. W. *Mar. Ecol. Prog. Ser.* 10, 57-65 (1982).
7. Richmond, R. H. *Mar. Ecol. Prog. Ser.* 22, 181-185 (1985).
8. Loya, Y. *Mar. Biol.* 13, 100-123 (1972).

Neutrality or selection?

In examining a large sample of parent-offspring data from two human major histocompatibility complex loci, *HLA-A* and *HLA-B*, from South Amerindians (F. L. Black and P. W. Hedrick, unpublished results), I have discovered a unique bi-allelic model. In this model there are no changes in allelic frequency for any given initial frequency — a neutral equilibrium — but there is a minimum mean fitness and a maximum excess of heterozygotes over Hardy-Weinberg expectations for equal allelic frequencies.

For all types of mating in the sample (see Table 1) that can produce homozygotes, except where the mother is a homozygote and the father is a heterozygote, there is a significant excess of heterozygotes over expected mendelian proportions, with a large (0.46) and statistically significant selection coefficient (s) against homozygotes¹. These results are suggestive of major histocompatibility complex (MHC) involvement in maternal-fetal interactions where homozygous offspring are selected against when carried by heterozygous mothers.

This selection is described by the bi-allelic model below, where alleles A_1 and A_2 have frequencies of p and q and genotypes A_1A_1 , A_1A_2 , and A_2A_1 have frequencies of P , H and Q . Note that Hardy-Weinberg proportions are not assumed and that only homozygous progeny from heterozygous mothers are selected against. The mean fitness is $\bar{w} = 1 - sH/2$ and the expected change in the frequency of A_1 is:

$$\Delta q = \frac{q(q - sH/2) + pq - q\bar{w}}{\bar{w}} = 0.$$

Quite surprisingly, there is no change in allelic frequency for any value of q (Fig. 1), just as for the neutrality model, the basis of molecular evolution theory, in which all genotypes have the same fitness.

The equilibrium frequency of heterozygotes is calculated by letting $H_e = H$ so that:

$$H_e = \frac{2pq}{(1 - sH_e/2)} = \frac{1 - (1 - 4spq)^{1/2}}{s}$$

The mean fitness at equilibrium is $\bar{w} = 1 - sH_e/2$. The equilibrium values of heterozygosity and mean fitness are reached quickly from any starting genotypic frequencies.

The mean fitness is a function of allelic frequencies with a minimum at $p = q$ and a maximum when $q = 0$ or 1 (Fig. 1), a pattern reminiscent of selection against heterozygotes². The fixation index $F = 1 - (H_e/2pq)$ is negative (Fig. 1), reminiscent of selection favouring hetero-

Table 1 Selection model

Parents		Progeny		
Female × Male	Frequency	A_1A_1	A_1A_2	A_2A_1
$A_1A_1 \times A_1A_1$	P^2	1	—	—
$A_1A_1 \times A_1A_2$	PH	1	1	—
$A_1A_1 \times A_2A_2$	PQ	—	1	—
$A_1A_2 \times A_1A_1$	PH	$1 - s$	1	—
$A_1A_2 \times A_1A_2$	H^2	$1 - s$	1	$1 - s$
$A_1A_2 \times A_2A_2$	HQ	—	1	$1 - s$
$A_2A_2 \times A_1A_1$	PQ	—	1	—
$A_2A_2 \times A_1A_2$	HQ	—	1	1
$A_2A_2 \times A_2A_2$	Q^2	—	—	1
		$\frac{p(p - sH/2)}{1 - sH/2}$	$\frac{2pq}{1 - sH/2}$	$\frac{q(q - sH/2)}{1 - sH/2}$

A bi-allelic selection model to describe observations of segregation for genes *HLA-A* and *HLA-B* in South Amerindians.

zygotes². The maximum excess of heterozygotes over Hardy-Weinberg proportions (the most negative F value) occurs when the allelic frequencies are equal. For example, when the allelic frequencies are equal and $s = 0.5$, then $\bar{w} = 0.854$ and $F = -0.172$.

Heterozygous females have lower fitness than homozygous females because of the lowered average fitness of their offspring. However, they appear to compensate precisely for this by the relatively higher fitness of offspring that are exactly like themselves (heterozygotes), thus explaining the neutrality equilibrium. The relatively higher fitness of their heterozygous offspring results in an overall excess of heterozygous offspring.

However, given two alleles in neutral equilibrium and the generation of a third allele by mutation, this new allele has a selective advantage. In other words, for situations with more than two alleles, such as most MHC genes, this model predicts that selection will maintain a stable polymorphism for multiple alleles.

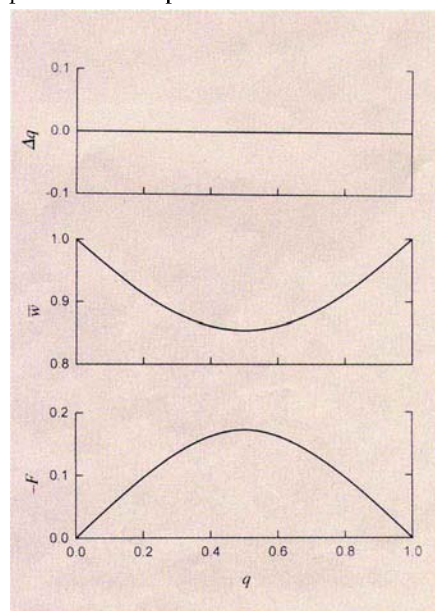


Figure 1 From top to bottom, the change in allelic frequency (Δq), the mean fitness ($\bar{w}$) and the excess of heterozygotes ($-F$) as a function of allelic frequency for the selection model when $s = 0.5$.

This bi-allelic selection model has no effect on the dynamics of genetic change, but when the two alleles have equal frequency the model results in a minimum fitness and a maximum excess of heterozygotes. This is, to my knowledge, the first description of a selection model with these unique properties.

Philip Hedrick

Department of Zoology,
Arizona State University,
Tempe, Arizona 85287, USA
e-mail: hedrick@hedricklab.la.asu.edu

1. Haldane, J. B. S. *J. Genet.* **54**, 294–296 (1956).
2. Hedrick, P. W. *Genetics of Populations* (Jones & Bartlett, Boston, 1985).

DNA answers the call of pipistrelle bat species

Groups of organisms that have been described as a single taxonomic unit on the basis of quantitative characters are increasingly proving to require more complex classification, when their evolutionary history is studied with molecular markers¹. Here we report an analysis of mitochondrial DNA sequences from the two echolocating types² of Europe's most abundant and well-studied bat, the pipistrelle³ (*Pipistrellus pipistrellus*). We describe genetic divergence that supports its reclassification as two different species.

Species-specific acoustic signals are especially important in nocturnal animals because of constraints on visual communication in the dark. Species can therefore often be identified by differences in their calls or other acoustic signals^{4,5} when morphology and behaviour are poor discriminative tools⁶. A bimodal distribution of echolocation call frequencies of the pipistrelle indicates that there may be a previously unrecognized taxonomic division².

Pipistrelles emit echolocation calls with the frequency of most energy close to either 45 or 55 kHz, yet the two phonic types are

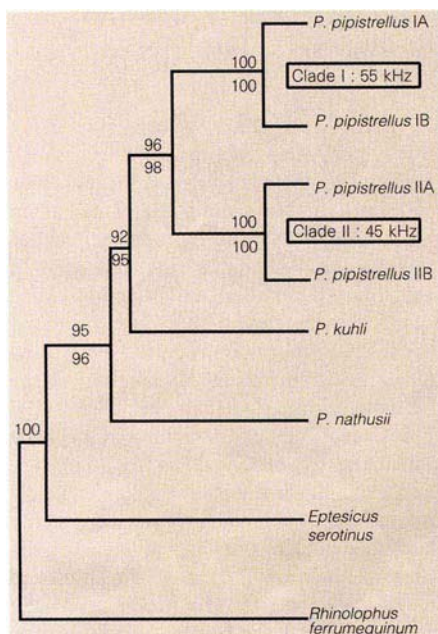


Figure 1 Phylogenetic tree for 630 base pairs of cytochrome *b*, showing the relationships between haplotypes and phonic types in *Pipistrellus pipistrellus* and related bat species (*P. nathusii*, *n*=2; all other genotypes, *n*=4). The tree represents the single topology found using maximum-likelihood, maximum-parsimony and neighbour-joining methods in PHYLIP¹³. Numbers indicate bootstrap values: above the branches from a neighbour-joining analysis, below the branches from a parsimony bootstrap analysis (1,000 replicates each). The separation of the two phonic types is fully supported, with a sequence divergence of 11%. There is up to 95% support for the separation of *P. pipistrellus* from the other species.

indistinguishable in multivariate analyses of flight morphology². Maternity colonies contain bats of only one phonic type, but the distributions of the two phonic types overlap for much of their European range².

We compared 630 base pairs of aligned cytochrome *b* sequences from four bats of each phonic group and other related species. This revealed four haplotypes which clustered into two distinct clades. Within each clade (termed I and II) the haplotypes showed a sequence divergence of less than 1%, whereas between clades the divergence exceeded 11% (Fig. 1) (NCBI accession numbers U95499–U95514). Divergence of this magnitude is similar to that found between reproductively and morphologically distinct bat species^{7,8}. Comparing the differences with data from the vertebrate molecular clock⁹, the phonic types of *P. pipistrellus* seem to have diverged 5–10 million years ago, and so could be considered as distinct phylogenetic species¹⁰.

Individuals from each maternity roost were of a single phonic type, which was associated with sequences from the same cytochrome *b* clade, but the abundance of

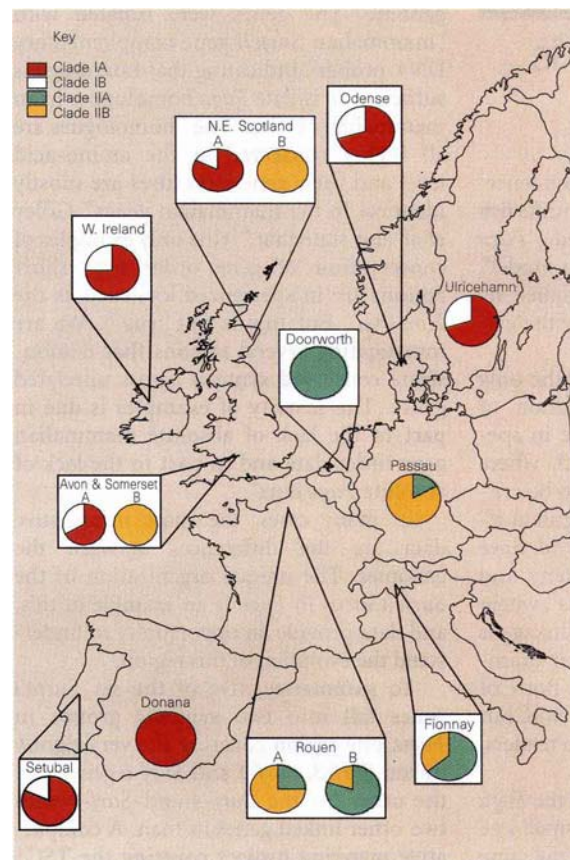


Figure 2 Distribution of mitochondrial DNA genotypes and phonic types of pipistrelle in Europe. Each pie chart represents one maternity roost. The colonies, with number of bats sequenced, echolocation call frequencies where available (and number of bats recorded) are: north-east Scotland (A) 12 sequenced; (B) 10 sequenced, 45.6±1.3 kHz (12 recorded); Avon and Somerset, UK (A) 15, 57.6±1.6 kHz (50); (B) 8, 46.1±1.5 kHz (39); Ireland (A) 8; Rouen, France (A) 8; (B) 5; Setubal, Portugal 11; Odense, Denmark 10; Ulricehamn, Sweden 23, 53.6±1.8 kHz (58); Passau, Germany 11; Fionnay, Switzerland 11; Doorworth, Holland 14; Donana, Spain 8.

the two haplotypes within each clade varied geographically (Fig. 2). In clade I (55 kHz), haplotype IB was always rarer. In clade II (45 kHz), British bats always belonged to haplotype IIB whereas roosts on mainland Europe often contained both haplotypes. The distributions of the two clades overlapped for much of Britain and mainland Europe, although clade I bats tended to occur near the edge of the pipistrelle's range (Ireland, Sweden and Portugal).

It is more likely that acoustic divergence between species occurs after genetic isolation, rather than functioning as a mechanism to promote speciation¹¹. In most animals (for example, grasshoppers, crickets, birds, frogs and toads) vocal divergence between species seems to be driven by female choice through sexual selection¹². For the echolocation calls of bats, however, natural selection seems to have shaped call design so that optimal echoes can be received from targets of different sizes. Bat echolocation signals may be more constrained than are mating songs of other animals, and are unlikely to become more elaborate as a result of sexual selection.

Our finding that one of Europe's most widespread, abundant and well-studied mammals exists as two cryptic species has implications for the study of bat biodiversity. If a mammal that has received considerable scientific attention has only now been shown to exist as two species, many of the over 900 bat species whose echolocation calls and genetics have yet to be studied may

also comprise groups of morphologically similar but evolutionarily distinct species.

**E. M. Barratt, R. Deaville
T. M. Burland, M. W. Bruford**

*Conservation Genetics Group,
Institute of Zoology, Zoological Society of London,
Regent's Park, London NW1 4RY, UK*

G. Jones

*School of Biological Sciences, University of Bristol,
Bristol BS8 1UG, UK*

P. A. Racey

*Department of Zoology, University of Aberdeen,
Aberdeen AB24 2TZ, UK*

R. K. Wayne

*Department of Biology, University of California,
Los Angeles, California 90024, USA*

1. Avise, J. C. *Molecular Markers, Natural History and Evolution* (Chapman & Hall, London, 1994).
2. Jones, G. & van Parijs, S. M. *Proc. R. Soc. Lond. B* **251**, 119–125 (1993).
3. Stebbings, R. E. & Griffith, F. *Distribution and Status of Bats in Europe* (NERC, Huntingdon, 1986).
4. Littlejohn, M. L. in *The Evolution of the Amphibian Auditory System* (eds Fritsch, B., Ryan, M. J., Wilczynski, W., Hetherington, T. E. & Walkowiak, W.) 613–630 (Wiley, New York, 1989).
5. Jones, G., Morton, M., Hughes, P. M. & Budden, R. M. *J. Zool. Lond.* **230**, 385–400 (1993).
6. Mayr, E. *Animal Species and Evolution* (Harvard Univ. Press, Cambridge, MA, 1963).
7. van den Bussche, R. A., Baker, R. J., Wichman, H. A. & Hamilton, M. J. *Mol. Biol. Evol.* **10**, 944–959 (1993).
8. Sudman, P. D., Barkley, L. J. & Hafner, M. S. *J. Mamm.* **75**, 365–377 (1994).
9. Moritz, C., Dowling, T. E. & Brown, W. M. *Annu. Rev. Ecol. Syst.* **18**, 269–292 (1987).
10. Cracraft, J. *Curr. Ornithol.* **1**, 159–187 (1983).
11. Butlin, R. K. *Am. Nat.* **130**, 461–464 (1987).
12. Andersson, M. *Sexual Selection* (Princeton Univ. Press, 1994).
13. Felsenstein, J. *PHYLIP (Phylogeny Inference Package)* (Department of Genetics, Univ. Washington, Seattle, 1993).

How good a model is the *Fugu* genome?

J. Gilley *et al.* in *Scientific Correspondence*¹ describe the organization of the pufferfish *Surfeit* locus in a letter entitled "*Fugu* genome is not a good mammalian model". The wide-ranging conclusions implied by this headline are not supported by the evidence and arguments presented.

First, Gilley *et al.* state that "[t]he only examples of significant conservation of gene order over short regions are in specialized loci such as the *Hox* loci, where genomic organization is believed to be critically related to developmental regulation". This is not true. M. K. Trower *et al.* have shown that conservation of synteny and gene order is seen in the *c-fos* locus², where there is no reason to believe that linkage is constrained by function, and other examples are provided by G. Elgar³. Both of these papers were cited by Gilley *et al.* but their contents were not brought to readers' attention.

Second, we originally proposed the *Fugu* genome as a model because of its small size and because we showed that it had the same overall gene repertoire as the human genome. Subsequent work has confirmed these properties and the proposed model. We recognized early on that where conservation of synteny and gene order were found, there would be additional advantages for comparative positional cloning, but we have been careful not to make large claims for this. Determination of the extent of linkage conservation will be important for our understanding of the evolution of the vertebrate genome and will be a matter of direct observation rather than opinion based on a single example.

Samuel Aparicio

Wellcome/CRC Institute,
Tennis Court Road,
Cambridge CB2 1QR, UK

Sydney Brenner

Molecular Sciences Institute,
9894 Genesee Avenue,
La Jolla,
California 92037, USA

Gilley and colleagues¹ present data on the *Surfeit* genes isolated from the *Fugu* genome. The *Surfeit* locus is of particular interest as the genes contained within are tightly packed in both mammals and the chicken and yet dispersed in invertebrates. Unfortunately, the data from *Fugu* do not support the conclusion drawn by Gilley *et al.* that *Fugu* is not a good mammalian genome model.

On the contrary, *Fugu* is a good model. All the *Surfeit* genes are present in the *Fugu*

genome. The genes were isolated with "mammalian *Surfeit* gene complementary DNA probes", indicating that homology is sufficient to isolate *Fugu* homologues from mammalian cDNAs. The "homologues are all highly conserved at the amino-acid level and their gene structures are mostly identical to the mammalian genes". Gilley *et al.* also state that "[t]he only examples of conservation of gene order over short regions are in specialized loci such as the *Hox* loci". But this is not true^{2,3}. We are investigating several regions that demonstrate conserved synteny across unrelated genes. The scarcity of examples is due in part to the lack of absolute mammalian gene order data and in part to the lack of absolute *Fugu* data⁴.

In many cases, the most informative data are the differences between the genomes. The unique organization of the *Surfeit* locus in *Fugu* is an example of this, and data provide an opportunity to understand the evolution of this region.

To summarize, five of the six *Surfeit* genes fall into two syntenic groups in *Fugu*; one region covering the very tightly linked *Surf-3*, *Surf-1* and *Surf-6* genes and the other linking *Surf-4* and *Surf-2* with two other linked genes in man. A comparative mapping project covering the TSC1 region on 9q34, using the *Fugu* genome, is under way to address the issue raised by the authors in a more comprehensive manner.

It is true that some regions of *Fugu* and mammalian genomes will bear no relation to each other and gene order will be fragmented. In the case of the *Surfeit* locus, there is undeniably a degree of fragmentation, but there is also enough conserved synteny to generate informative data. It is unfortunate that Gilley *et al.* have created an impression that comparative genomics with *Fugu*, and perhaps with lower organisms generally, is not important — this is simply not the case.

Greg Elgar

Melody Clark

UK HGMP-Resource Centre,
Hinxtion Hall,
Cambridge CB10 1SB, UK
e-mail: gelgar@hgmp.mrc.ac.uk

Andrew Green*

Richard Sandford†

*Department of Medical Genetics,
and †Department of Medicine,
University of Cambridge,
Addenbrookes NHS Trust,
Hills Road,
Cambridge CB2 2QQ, UK

1. Gilley, J., Armes, N. & Fried, M. *Nature* **385**, 305–306 (1997).
2. Trower, M. K. *et al.* *Proc. Natl Acad. Sci. USA* **93**, 1366–1369 (1996).
3. Elgar, G. *et al.* *Trends Genet.* **12**, 145–150 (1996).
4. How, G.-F., Venkatesh, B. & Brenner, S. *Genome Res.* **6**, 1185–1191 (1997).

Aaaargh, no! – in lights

Using the grisly example of a tiger-eaten tour guide, and the more subtle alarm pheromone of a pike-eaten minnow, Jared Diamond¹ in *News and Views* discussed recent work² identifying one benefit of alarm calls — predator interference. In addition to the audible and olfactory alarm signals that he noted, there are the many visible ones, most notably the flashes produced by luminescent organisms finding themselves similarly *in extremis*.

These visible signals resulted in the 'burglar alarm' hypothesis, whereby secondary predators, either different species or larger conspecific cannibals, are alerted by the flash to potential prey in the form of the primary predator or grazer³. The firefly in the spider's web, the dinoflagellate in the copepod's grasp, the brittle star grabbed by the crab and the ctenophore bitten by the turtle all 'scream' visibly. The value of these responses (and the validation of the hypothesis) has recently been demonstrated by experiments in which non-luminous prey, swimming in the dark among luminous dinoflagellates, were directly targeted by secondary visual predators (both fish and squid) using the flashes of the dinoflagellates^{4,5}.

Most of the oceanic environment is permanently dark (and much of the rest of it is dark during the night). A majority of marine individuals (and, in many taxonomic groups, most species) are luminous. Visible signals are probably the commonest alarm response in this habitat. Their success depends on the likelihood of a secondary predator within visible range, probably no more than a few tens of metres. Population densities decline logarithmically with depth, so primal flashing should be less effective at great depth, though it might still persist if it has a *direct* deterrent effect on the primary predator. Copepods eat fewer luminous dinoflagellates than non-luminous ones at similar densities, so the two effects can go hand in hand^{6,7}.

I do not know whether a minnow alarm pheromone intimidates a pike, or whether a scream deters a tiger, but in neither individual case is the *mechanism* of the effect relevant, only its time constant.

Peter J. Herring

Southampton Oceanography Centre,
Empress Dock, Southampton SO14 3ZH, UK
e-mail: p.j.herring@soc.soton.ac.uk

1. Diamond, J. *Nature* **385**, 295–296 (1997).
2. Chivers, D., Brown, G. & Smith, J. *Am. Nat.* **148**, 649–659 (1996).
3. Burkenroad, M. D. *J. Mar. Res.* **5**, 161–164 (1943).
4. Mensinger, A. F. & Case, J. F. *Mar. Biol.* **112**, 207–210 (1992).
5. Fleisher, K. J. & Case, J. F. *Biol. Bull.* **189**, 263–271 (1995).
6. Esaias, W. E. & Curl, H. C. *Limnol. Oceanogr.* **17**, 901–906 (1972).
7. White, H. H. *J. Exp. Mar. Biol. Ecol.* **36**, 217–224 (1979).

From Plato to Penrose

Insights of Genius: Imagery and Creativity in Science and Art

by Arthur I. Miller

Springer-Verlag: 1996. Pp. 482. £15.95, \$27

John Mollon

This is a ramble of a book. If there is a central theme, it is the role of visual representations in scientific discovery and theory. But in effect the 480 pages are devoted to a long nature walk through the history of physics. The walk takes us to and fro between the centuries and often — too often — returns us to the author's favourite period, 1905–27.

Along the way our author-guide offers brief biographies of Galileo, Poincaré, Einstein and Picasso, warmly attacks the post-modernists who maintain that science is a mere social construct, describes scientific progress in terms of the Piagetian concepts of assimilation and accommodation, and assesses whether Picasso was influenced by Poincaré's writings on geometry. Everyone from Kant to Kosslyn, from Plato to Penrose, has a paragraph.

We learn that the first non-Euclidean geometry was published in the *Kazan Messenger*, that Poincaré's examiners judged his doctoral thesis confused, that Picasso said of Braque "*C'était comme si nous étions mariés*," ["It was as if we were married"] and that Mondrian resigned from the "*De Stijl*" movement when another member, his fellow Dutch painter van Doesburg, advocated the use of diagonal lines.

Grand issues there certainly are among the anecdotes — issues at the heart of nuclear physics, epistemology, aesthetics, cognitive science and neurophysiology. But every time the reader feels that he has grasped Ariadne's thread, the author makes a knight's move through the undergrowth and is in a different century discussing a different concept — with different metaphors.

The central theme — the role of visual representations — deserves a more focused treatment. Arthur Miller writes as if the issue has not been discussed before, whereas of course it has; notably missing from his bibliography are Hans Reichenbach's *Philosophie der Raum-Zeit-Lehre* (de Gryuter, 1928) and Simon Altmann's *Icons and Symmetries* (Clarendon Press, 1992). And, curiously, Miller says nothing about the history and role of graphs. (Does the frequency of tables, and rarity of graphs, in nineteenth-century journals reflect only the cost of printing blocks or does it tell us that modes of thought were different?)

Miller rightly takes as instructive the transition from Niels Bohr's 1913 Solar-



System model of the atom to Werner Heisenberg's quantum mechanics. He distinguishes between *Anschaung* (visualization) and *Anschaulichkeit* (visualizability): in the former case the imagery of a scientific model is derived from phenomena we have witnessed in the macroscopic world; in the latter, the imagery reflects the structure of the scientific theory. After 1927, physicists had to accept that they could not directly visualize the subatomic world. (This is reasonable: our brains evolved to represent a macroscopic, Newtonian world.) But visual imagery could be used to represent the mathematics, notably by means of Feynman diagrams.

It is interesting to compare Miller's and Altmann's treatments of Hans Christian Ørsted's discovery of 1820 that a compass needle can be deflected by electric current passing through a wire. Miller gets the date badly wrong and presents the discovery as one coming easily to a mind prepared by *Naturphilosophie*: Ørsted believed in a universal vital force and so expected to find an interaction between electricity and magnetism. Altmann, in contrast, suggests that Ørsted came upon the effect (perhaps accidentally) only after eight years of trying the wrong experiment, eight years of avoiding the critical arrangement in which the platinum wire is parallel to the meridian plane in which the compass needle rests.

Altmann believes Ørsted was in fact held back by a visual representation, one in which both the magnetic and the electric vectors were represented as polar. In such a model, the current acts symmetrically upon the parallel needle and no deflection is expected.

Miller has a good discussion of the status of visual imagery in experimental psychology. He rehearses the debate between Stephen Kosslyn and Zenon Pylyshyn about whether images play a causal role or are epiphenomena and whether spatio-visual information is represented within our brain propositionally (as in digital computers) or in an analogue way. He evaluates shrewdly the studies using positron emission tomography that show increased blood flow in visual parts of the brain when the subject is asked to manipulate remembered spatial information.

He passes on to Einstein's thought experiments of 1895 and 1907 and then writes lamely: "These conclusions go far to substantiate that visual images are not epiphenomena and are essential to scientific research." Now, it is plausible to suppose that Einstein did use the visual parts of his brain in developing both special and general relativity. But the fact is that we understand neither the nature of consciousness nor the format of visual representations in the brain. Einstein's insights do not bear one way or another on whether images are epiphenomena or on the other issues that Miller discusses earlier in the chapter.

Insights of Genius could usefully have been tighter and shorter. But the next time I am asked for vacation reading by an undergraduate who has a taste for grand ideas, I shall recommend this book as an introduction to the glorious jumble that is the history of Western thought. □

John Mollon is in the Department of Experimental Psychology, Cambridge University, Downing Street, Cambridge CB2 3EB, UK.

Star-gazing giant

Yerkes Observatory, 1892–1950: The Birth, Near Death, and Resurrection of a Scientific Research Institution

by Donald E. Osterbrock

University of Chicago Press: 1997. Pp. 384.

£31.95, \$40

Jack Meadows

One enduring astronomical debate relates to the siting of observatories. On the one hand, they need clear skies; on the other, they must be accessible. As transport and communication facilities have improved, so sites have increasingly been chosen for the clarity of the night sky.

Yerkes Observatory was born towards the end of the nineteenth century at a crucial moment in this retreat from human habitation. When George Hale at the University of Chicago persuaded the millionaire Charles Yerkes to donate money to buy the largest telescope in the world, the university expected that the observatory would be built in the city. Persuasion by the astronomical community led to its erection out in the country, but at a low-lying site by Williams Bay in Wisconsin.

That period was also a time of transition in telescope design. The most favoured type of telescope in the nineteenth century had been the refractor. Yerkes' largest instrument was, correspondingly, a huge refracting telescope with a main lens 40 inches in diameter. But opinion was already coming around to the view that large telescopes in the future would be reflectors, using mirrors rather than lenses. So Yerkes occupies a transitional position between the nineteenth and the twentieth centuries as regards both site and equipment.

Donald Osterbrock's book covers the history of the observatory under its first three directors. A major theme is how they variously coped with the astronomical environment that the site and equipment provided. The first director, Hale, did not stay long. By 1904, he had migrated to California, where he subsequently erected a large reflecting telescope on the top of Mount Wilson, correctly foreseeing the way in which observational astronomy would develop in the twentieth century.

Hale left Edwin Frost as his successor at Yerkes. As Osterbrock shows well, Frost was a charming man, but had little of Hale's pioneering spirit. The research activities of the observatory slid slowly downhill during the three decades of his reign.

Frost's successor, Otto Struve, a Russian émigré from a distinguished astronomical family, was much more in the Hale mould. He developed joint activities with the new McDonald Observatory in Texas which allowed staff at Yerkes regular access to a large reflecting telescope on a mountain top. This proved a good way of overcoming the limitations of the Yerkes site.

More importantly, Struve pressurized the University of Chicago into supporting a remarkable mix of outstanding theoretical and observational astrophysicists. For a time, immediately before and after the Second World War, Yerkes became one of the most exciting places in the world for astrophysical research.

These central chapters of the book are among the most interesting. Osterbrock's extensive use of archival material makes plain how difficult it was for such a group of prima donnas to remain in harness together. Struve's choice of staff was triumphantly vindicated — Chandrasekhar, for example, became one of the first astrophysicists to win

a Nobel prize. But, when increasing tension led to Struve's departure from Yerkes in 1950, it was a sign that the golden years were over. Although good work remained to be done, the observatory has never reached the same heights since.

This is an excellent description of the ups and downs of a major observatory, told by someone who himself directed another leading US observatory — Lick in California. My only criticism is that few of the stories with which Yerkes staff entertained visitors are mentioned. What about the gargoyle of Charles Yerkes that is supposed to have been carved on the building, or the photographic plate with a bullet-hole through it from the Chicago gang wars? □

Jack Meadows is in the Department of Information and Library Studies, Loughborough University of Technology, Loughborough, Leicestershire LE11 3TU, UK.

The erotic ape

Bonobo: The Forgotten Ape

by Frans de Waal and Frans Lanting

University of California Press: 1997. Pp. 210.

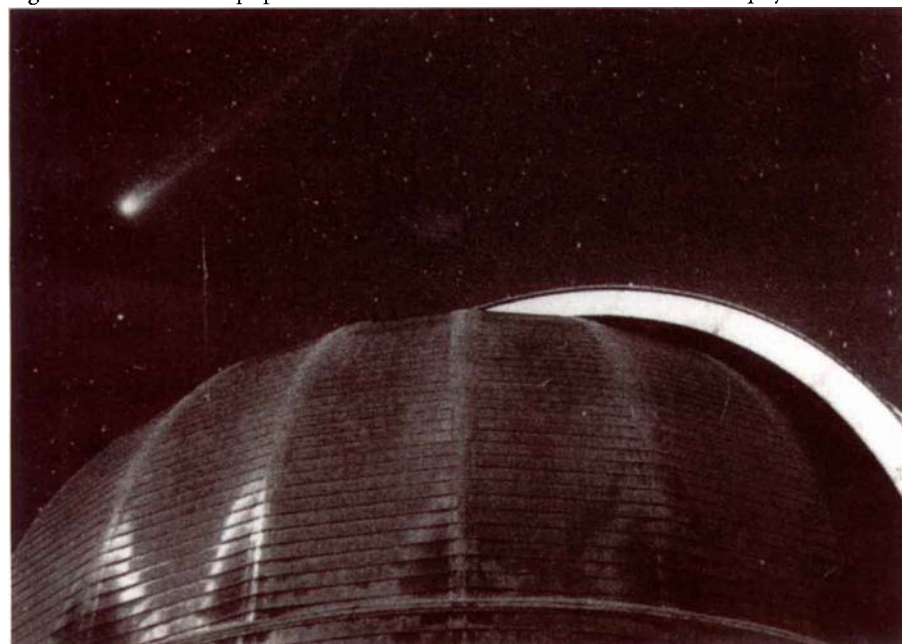
\$39.95

William McGrew

Apart from ourselves, there are only four living species of great apes, and three of these are well-known to Western science, largely through the efforts of Louis Leakey's protégées: Jane Goodall on chimpanzees (*Pan troglodytes*), Dian Fossey on mountain gorillas (*Gorilla gorilla*) and Biruté Galdikas on orangutans (*Pongo pygmaeus*). The fourth species, *Pan paniscus*, the bonobo or pygmy chimpanzee, remains relatively obscure. It is arguably better known to Eastern science, with two of the previous three books on the species coming from Japanese primatologists (Takayoshi Kano, *The Last Ape*, 1986 & 1992; Suehisa Kuroda, *The Unknown Ape*, 1982).

Now, finally, comes a synthesis of what is known about bonobo behaviour in captivity and in nature, by a pair of Dutchmen, laboratory ethologist Frans de Waal and wildlife photographer Frans Lanting. De Waal provided the text and notes, while Lanting journeyed to Zaïre to photograph the apes of Wamba. The result is a handsome, semi-coffee table volume with more than 70 colour plates, yet with an extensive bibliography and nine pages of detailed notes. The text is crisp and clean, enlivened by de Waal's gently ironic humour, which readers of his previous books, *Chimpanzee Politics* (1982), *Peacemaking among Primates* (1989) and *Good Natured* (1996), have come to expect.

The bonobo is not so much the forgotten ape as the ape yet to be studied. Discovered by science little more than 60 years ago, the species lives naturally only in remote areas of



Country air, clear skies: comet Hyakutake over the Yerkes Great Dome.

one equatorial African country. That country, Zaïre, is one of the world's most troubled, and at present all field research on the species is suspended. In captivity worldwide, there are barely 100 bonobos, mostly in zoos; in contrast, there are about 1,800 chimpanzees in American laboratories alone. All of these factors conspire to limit our knowledge of the fourth great ape to a fraction of what is known of the other three species, despite heroic efforts by primatologists.

Given this disparity in knowledge, it is not surprising that the pervasive theme of the book is comparison: How closely does the bonobo resemble its congener, the chimpanzee? How different are bonobos in captivity from those in nature? To what extent are bonobos unique among the apes? Not surprisingly, the answers vary by topic, according to the quality and quantity of available evidence.

De Waal's main interests are in social relations, so he focuses on the one field site, Wamba, from which by far the most data on social behaviour come, thanks to the efforts of Takayoshi Kano, the doyen of wild bonobo research, and his team. Unfortunately, the findings from Wamba are confounded by methodology — the apes there continue to be fed by scientists with large amounts of highly prized domesticated plant foods. Several of Lanting's photographs show the ground in the artificial feeding area to be strewn with sugar cane. Such massive dietary supplementation with high-calorie food distorts many aspects of behaviour, from group ranging to individual bipedality, and so it has been abandoned or avoided in virtually every other study of African apes.

This matters because some of the differences across species or study sites may be due to this feeding, not to ecological or cultural differences between populations. The bonobos of Wamba go about in big, sociable parties, for example, while their counter-



Secret ape: little is known about bonobo behaviour.



Sexual freedom: "it is unlikely that a male bonobo can distinguish his offspring from those of other males".

parts at Lomake, a site where food has never been provided by humans, travel in small groups no different in size from those of chimpanzees. So a question mark must be attached to many of the points in de Waal's arguments on the past evolution and present function of bonobo social organization, as these are compromised by the field data upon which he depends.

The authors might have called the book *The Erotic Ape*, for the rich and varied expression of bonobo sexuality is found wherever the species is studied. Here, truly, is an organism that practises recreative rather than procreative sex, with little regard to constraints of gender, age or rank, although the line is drawn at incest. This is doubly impressive because of its social, liberated nature. There is no conclusive evidence that bonobo reproductive physiology is any different from that of chimpanzees, despite claims to the contrary. Further, although ejaculation is limited to heterosexual coitus, stimulation to orgasm seems open to any combination of participants. 'Foreplay' seems to be extended to before, during or after any socially arousing context.

In other aspects of daily life, whether or not bonobos differ notably from chimpanzees is largely a matter of 'is the glass half-empty or half-full?' Unlike chimpanzees, bonobos lack subsistence technology, infanticide, male-male coalitions, and male-dominated access to resources. Some aspects of their social life, such as adult male reliance upon social support from their mothers, seem to be unique among all primates. But most features, such as fission-fusion social structure, female bonding, social intelligence, feeding ecology, sexual dimorphism, bipedality and arboreality, seem to be variations on the basic *Pan* theme.

The authors do not ignore issues of con-

servation. None of the sites where wild bonobos have been studied is legally protected, and their future must be assumed to be precarious, especially because of the demands of the bush meat trade. In captivity, the small bonobo population is spread over too many places. The median colony size is only five individuals, which is far too few for the complex sociosexual lives of this fascinating species to be played out. Perhaps the impact of this milestone book will help such matters.

William McGrew is in the Departments of Anthropology and Zoology at Miami University, Oxford, Ohio 45056, USA.

The reality of relativity

Einstein's Mirror

by Tony Hey and Patrick Walters
Cambridge University Press: 1997. Pp. 287.
£50, \$69.95 (hbk), £16.95, \$27.95 (pbk)

Frank Close

The two great foundations of twentieth-century physics are quantum theory and relativity. In 1987, Tony Hey and Patrick Walters wrote *The Quantum Universe*, a liberally illustrated and highly successful book showing how the arcane quantum theory affects our daily experience. They have now followed the format of that book and taken on relativity, in a popular exposition that will be high on the reading lists of university and advanced high-school students and general readers with an interest in science.

Whereas quantum theory has many strands and inventors — including Albert Einstein, whose Nobel prize was for the photoelectric effect — relativity and Einstein are synonymous. Einstein has already been the subject of so many biographies and commentaries that when a new book about

his work appears, a common reaction is likely to be "What's new?" The answer, in the case of *Einstein's Mirror*, is its emphasis on the way that relativity theory is rooted in experiment and, as in the authors' earlier book on the quantum theory, how it is manifested in nature.

There is a widespread misapprehension in the public mind that Einstein overthrew the existing paradigms by thought alone. There are many enthusiastic individuals who, inspired by the Einstein myth, believe that they too have tapped the secrets of the Universe by their unique insight, without need or care for experiment. Although Einstein is best known as the creator of relativity, it is less widely realized that he spent most of his life in a fruitless quest for a unified theory of gravity and electromagnetism. Hey and Walters recall Richard Feynman's comment that Einstein's great works had sprung from physical intuition and that it was "when he stopped thinking in concrete physical images and became a manipulator of equations" that Einstein's creativity ceased. *Einstein's Mirror* shows the physical reality of relativity and how it relates to our daily lives, often in unexpected ways.

The title of the book recalls how the young Einstein was haunted by a conundrum: would his image appear in a mirror if both he and the mirror were moving at the speed of light? Einstein solved this mirror problem with his creation of special relativity; although he was not the only scientist

concerned with the problem of how the velocity of light might be affected by the motion of the observer, it was "Einstein alone who realized that the solution to the problem involved a radical rethinking of the concept of time". His new vision of space and time has routine applications in physics, chemistry and medicine — positron emission tomography, satellite navigation systems, particle accelerators, the famous $E=mc^2$, nuclear power and more are covered in the book.

The "happiest thought" of Einstein's life was that when someone is in free fall there exists, in their immediate surroundings, no gravitational field: "If the observer drops some bodies then these remain to him in a state of rest". Einstein had realized that in free fall, gravity is effectively abolished and his ideas on special relativity could be applied. Conversely, just as gravity can be 'abolished' by free fall, an artificial gravity can be created by acceleration: "Gravity and acceleration are, in a certain sense, equivalent".

This eventually led Einstein to his general theory of relativity, with its bizarre implications for the slowing of time in gravitational fields. (Incidentally, the time-dilation effects of motion in high-flying aircraft and the gravitational effects of the Earth almost cancel out.) These effects have been measured on Earth with atomic clocks, but they become much greater on neutron stars, and singular in black holes. In a section called 'Life on a Neutron Star', we learn that office workers on high floors "would live shorter lives than ground floor

workers although they would accomplish the same amount of work in a lifetime".

In Newtonian theory, gravity acts instantaneously, whereas according to Einstein, disturbances in the gravitational field are limited by the speed of light. This implies the existence of gravitational radiation, which has not yet been detected directly, although in 1978 Russell Hulse and Joseph Taylor found indirect evidence for it in the motion of a pulsar around a companion star. According to theory, the rotating binary system should emit gravitational radiation, although it would be too weak to measure directly. However, this gravitational radiation of energy shows up as a tiny decrease in the orbital periods of the pulsar, about one-tenth of a second in a year.

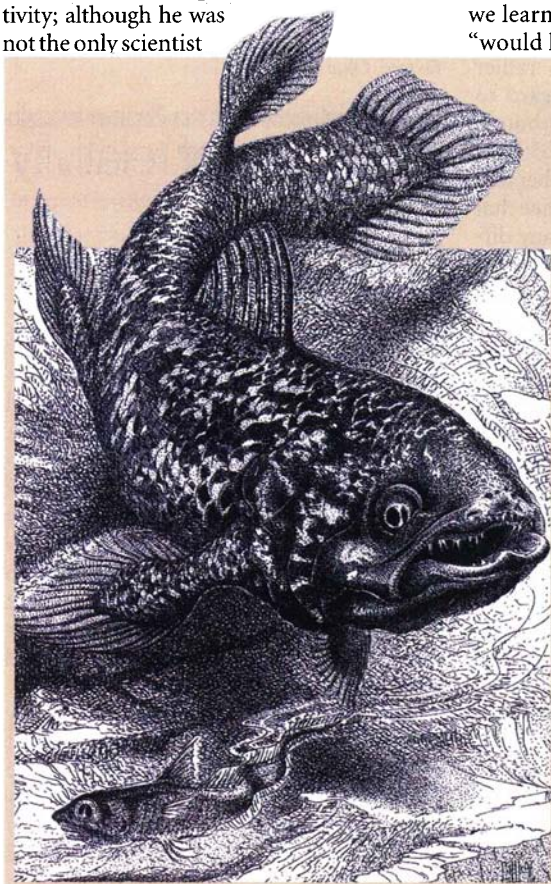
Fifty years ago, our vision was limited to the rainbow; then we expanded our electromagnetic senses into radio, infrared, ultraviolet, X-ray and gamma-ray astronomies, which have led to marvels and monsters that were undreamed of. By the time of the centenary of the theory of general relativity in 2015, we are likely to be exploring the dark side of the Universe by means of gravitational-wave astronomy. If history is any guide, the resulting discoveries will be stranger than fiction. The greatest legacies of Einstein's mirror may be yet to come, but in the meantime *Einstein's Mirror* is an excellent introduction to the implications and applications of relativity. □

Frank Close is in the Rutherford-Appleton Laboratory, Chilton, Didcot OX11 0QX, UK.

New Journals

This year, *Nature's* annual new journals review supplement will appear in the issue of 11 September. Publishers and learned societies are invited to submit journals for review, as well as details of any eligible electronic journals, taking note of the following criteria:

- Journals that first appeared during or after June 1995 and issued at least four separate numbers by the end of May 1997 will be considered.
 - Journals covering any aspect of science are eligible, although those dealing with clinical medicine and pure mathematics are excluded, as are publications of abstracts.
 - Frequency of publication must be at least three times a year. The main language used must be English. Translation journals in English are, of course, eligible.
 - Deadline for submission is 6 June.
- When submitting journals for review, please send at least four different issues (the first, the most recent and any two others) of each title, together with full details of subscription rates. For further information please contact Peter Tallack, *Nature*, Macmillan Magazines, Porters South, Crinan Street, London N1 9XW, UK. Tel: +44 (0)171 843 4567. e-mail: p.tallack@nature.com.



Exploring Earth's last frontier

The coelacanth was once thought to have accompanied the dinosaurs to extinction but stunned the scientific world in 1938 by turning out to be very much alive. There could be other 'living fossils' waiting to be discovered in the depths of the oceans — along with lost treasures and precious metals. William J. Broad uncovers some of these mysteries in *The Universe Below: Discovering the Secrets of the Deep Sea* (Simon & Schuster, \$30). Broad draws on extensive fieldwork and many interviews with scientists to explain how the explorers of the oceans go about their work using once-secret technologies developed during the Cold War. The book is written very much from an American perspective. The Pulitzer prize-winning author writes about science for *The New York Times*. Dimitry Schidlovsky provided the illustrations.

Seismic evidence for small-scale heterogeneity throughout the Earth's mantle

Michael A. H. Hedlin, Peter M. Shearer & Paul S. Earle

Institute of Geophysics and Planetary Physics, University of California, San Diego, La Jolla, California 92093-0225, USA

Details in the amplitude structure of short-period precursors to the seismic core phase PKP can be used to constrain the depth extent of mantle scattering. The simplest model consistent with the observations invokes small (~8 km), weak (r.m.s. velocity perturbations of 1%), random heterogeneities uniformly distributed throughout the mantle. The data do not support previously proposed models that place an increase in the concentration of scattering sites near the base of the mantle, and instead place an upper limit on the amplitude of any short-wavelength topography at the core-mantle boundary.

Seismic energy scattered from small-scale wave-velocity perturbations in the Earth generally arrives in the coda following the major seismic phases as the ray paths of the scattered waves have longer travel times than the direct arrivals. Possible weak scattering from deep structures is normally obscured by the much stronger scattering that results from near-surface crustal scattering and reverberation. However, the sharp velocity decrease at the core-mantle boundary (CMB) creates unusual ray geometries in which deep scattered energy can arrive before any of the major phases, providing a unique window into the small-scale structure of the deep Earth. The resulting faint short-period precursors to the earliest branch of PKP were first observed by Gutenberg and Richter¹, and there has been a continuing debate about their origin and what they imply about the Earth's velocity structure. Early hypotheses included diffraction of PKP at the CMB², refraction in the inner core³, and reflection or refraction of PKP at transition layers between the inner and outer cores^{4,5}. But it is now generally accepted that the precursors result from scattering of PKP off volumetric heterogeneities in the lowermost mantle and/or topography on the CMB^{6,7}, and this interpretation has been supported by analyses of the travel times, amplitudes, slownesses, and frequency content of precursors recorded by arrays of seismometers⁸⁻¹¹.

Researchers have modelled the amplitude and range dependence of the precursors by assuming statistical models of small-scale velocity heterogeneity. Although some of the older studies suggested that the scattering region extends at least 600–900 km above the CMB with velocity perturbations of a few per cent of more^{11,12}, other analyses indicated that the data could be explained with much thinner layers within the D'' region at the base of the mantle¹³ or by short-wavelength CMB topography¹⁴. Recent studies have focused on the CMB region as the source of the scattering, suggesting that the data could be explained equally well by either 0.5–1.0% r.m.s. velocity perturbations in a 200-km-thick layer at the base of the mantle, or by CMB topography with r.m.s. height of about 300 m (refs 15, 16).

Resolving the location and nature of the scatterers is important because they provide some of the best constraints on small-scale heterogeneity in the deep Earth. But several significant issues remain unresolved, including the relative importance of topographic versus volume scattering, the scale and strength of the irregularities, the depth extent of any mantle heterogeneity and the possibility of lateral variations in the scattering strength. Here we present a detailed image of the time- and range-dependence of PKP precursor amplitudes, obtained by stacking 1,600 global seismo-

grams. Modelling of these observations suggests that the precursors are caused by scattering from small-scale heterogeneity throughout most of the mantle. Some scattering from the CMB region is required to explain the onset of the precursors, but the CMB scattering appears no stronger than that occurring in the rest of the mantle. Velocity perturbations are about 1% at correlation scale lengths of 8 km. These results support models of small-scale compositional heterogeneity in the mantle¹⁷⁻¹⁹.

PKP precursors

Figure 1 shows ray paths and travel-time curves for PKP and the region for which precursors may be observed. As hypothesized by Haddon⁶, the earliest-arriving inner core branch (PKP_{df}) plays no role in the generation of these precursors but simply provides a reference phase for the earliest scattered waves to precede. The precursive arrivals stem from scattering off irregularities at the CMB or within the mantle, which deflect the ray paths of the mantle phase P to the outer-core phases PKP_{ab} and PKP_{bc}. Scattered energy is able to precede the df branch because of the unusual ray geometries arising from the sharp drop in P-wave velocity at the CMB. The earliest possible arrivals at any range are caused by scattering at the CMB; energy scattered above the CMB must arrive later.

Figure 2 shows examples of PKP precursors observed on individual seismograms between 118° and 145° range. The traces show increasing precursor amplitudes with increasing range. The time spanned by the precursor wavetrains reaches a maximum of roughly 18 s at 130° range. By 144.6° there is no evidence of any signal preceding PKP_{df}. The df coda amplitudes vary in a similar fashion, increasing with increasing range until the final trace at 144.6°.

Global average precursor amplitudes

From the on-line waveform archive of the Incorporated Research Institutions in Seismology (IRIS), we selected 1,600 broadband recordings of precursors made at source-receiver ranges between 118° and 145°. These traces are culled from a much larger set by using only those traces that exhibit stationary and low levels of noise. No consideration is made of the apparent presence or absence of precursors to avoid biasing our estimates of average precursor amplitudes. Next, we bandpass-filter each trace, compute its envelope function, and adjust its amplitude downward by an amount determined by the average energy level of the noise. Finally we align the records on the PKP_{df} arrival, correct the range for the effect of variable source depth, and stack the envelope functions in 1.5-s time bins and 1° range bins, with a weighting that gives preference to

those traces with the highest signal-to-noise ratios. We estimate error bounds on the stack using a bootstrap technique²⁰ in which we repeat the stack many times for random samples of our data.

The resulting stack, $S(r, t)$, shown in Figs 3 and 4, reveals a clear increase in precursor amplitude with increasing range and with increasing time. The correspondence between the onset of energy and the minimum travel-time curve for CMB scattering agrees with previous results⁷, and suggests that some contributions are coming from the depth of the CMB. It seems likely that the signal onset continues to track the CMB scattering curve below 125°, but this is difficult to resolve because of low precursor energy levels.

Modelling precursor amplitudes

Our primary goal is to model the amplitudes in this stack to determine the depth range of the scatterers that contribute to the PKP_{df} precursors. Factors that will affect precursor amplitudes include the intrinsic strength and scale of the heterogeneities, the

volume of scattering, the amplitude of the primary wave and the scattering angle.

With ray-tracing we can identify the places in the Earth from which any point in the stack could have received scattered energy (assuming the Earth reference model PREM and single scattering). As an example of this (Fig. 5) we show scattering isotimes for PKP_{bc} for a grid of scatterers located on the CMB near the receiver with the energy being recorded at two stations (located 130° and 140° from the source). All the scatterers located on the “–5 s” isotime contours in the upper and lower left panels of this figure will contribute energy to the stack at $S(140, -5)$ and $S(130, -5)$ respectively.

Figure 5 demonstrates that a unique, one-to-one, association between points in the stack, $S(r, t)$, and specific scatterers within the Earth cannot be made. At any range the energy which is associated with the most limited ambiguity is that which arrives first. As shown by the curves in Fig. 3, the earliest possible arrivals must have been

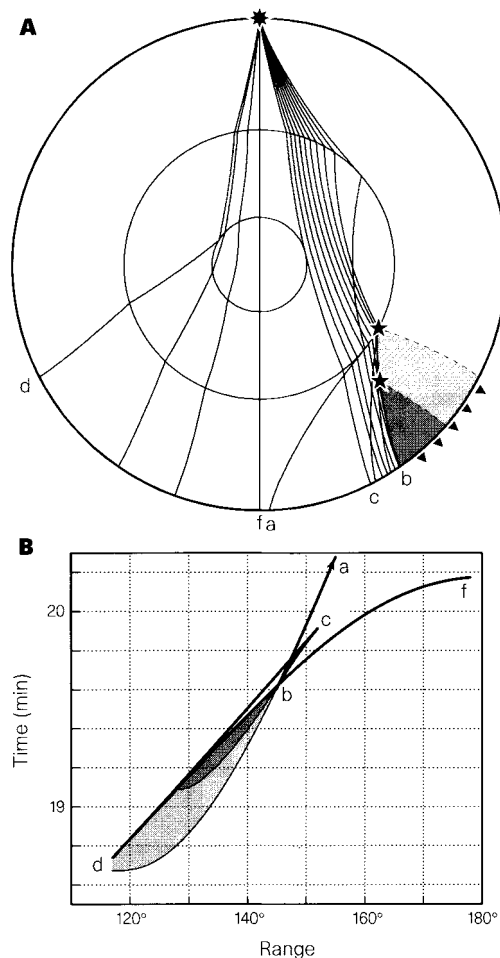


Figure 1 **A**, Ray paths of the different branches of PKP. The ab branch turns near the middle of the outer core, the bc branch turns just above the inner core, the cd branch is reflected from the inner-core boundary and the df branch turns within the inner core. Two scatterers (denoted by five-pointed stars) on the CMB and in the mantle near the receiver give rise to precursors that traverse the shaded regions. Although we have depicted PKP to P scattering into the shaded region on the receiver side of the travel path, precursors can also result from scattering on the source side (that is, from P to PKP). **B**, Travel-time curves for PKP are shown as heavy lines. The PKP precursors are observed to precede PKP_{df} by up to ~20 s at ranges less than the B caustic at 145° and are enriched in high frequencies. Energy scattered at the CMB can arrive as precursors to PKP in the lightly shaded region, scattered arrivals from a point in the mantle above the CMB can arrive in the darkly shaded area. Most of the scattered energy originates from the ab and bc branches near the B caustic.

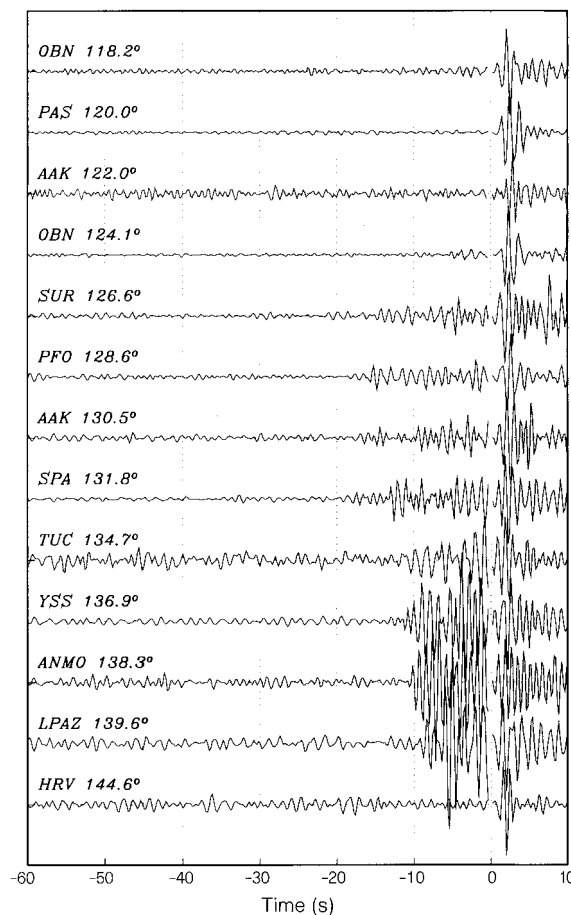


Figure 2 IRIS broadband recordings of PKP_{df} and preceding noise and precursors at ranges between 118° and 145°. These traces have been bandpass-filtered between 0.7 and 2.5 Hz and aligned on the PKP_{df} arrival. Amplitudes at negative times have been magnified by a factor of 7. The precursors are first visible in the OBN recording at 124.1° and last seen in the LPAZ recording at 139.6°. They become more energetic with increasing range. The earliest precursors precede PKP_{df} by 18 s (PFO at 128.6°). In some traces (for example, AAK at 122.0°) faint precursors may be obscured by pre-event noise.

scattered from the depth of the CMB. Later arrivals are associated with progressively more ambiguity as to the depth of scattering, because they represent a sum of energy scattered over a depth interval between the CMB and a shallower depth.

These ray-tracing results can also be used to compute the scattering angles (the angle by which the energy was diverted from its original path). In Fig. 5 we plot the PKP_{bc} scattering angles for CMB scatterers at 130° and 140° range. The scattering angle increases strongly with decreasing range, and also increases with increasing time and with the distance of the scatterer above the CMB. The increase in precursor amplitudes versus range is primarily due to the change in scattering angle, with stronger scattering occurring at smaller angles.

To model our observations, we assume that the precursors are due to the propagation of PKP through a random, inhomogeneous medium and use Chernov's acoustic scattering theory²¹ adapted to elastic media¹³. To simplify the analysis we assume that the mantle behaves like a Poisson solid, that the random perturbations in velocity have an exponential autocorrelation function²², and that we can neglect the influence of multiple scattering²³. With these assumptions, the average scattered power is given by:

$$\langle |\Phi_s(\theta)|^2 \rangle = \frac{2k^4 a^3 \bar{\alpha}^2 V A^2}{\pi r^2} \frac{\frac{1}{4} [\cos(\theta) + \frac{1}{3} + \frac{2}{3} \cos^2(\theta)]^2}{[1 + 4k^2 a^2 \sin^2(\theta/2)]^2} \quad (1)$$

where θ is the scattering angle, A is the incident wave amplitude, V is the volume of scattering, $\bar{\alpha}$ is the r.m.s. velocity perturbation ($\delta\alpha/\alpha_0$), k is the wavenumber (ω/α_0), a is the characteristic scale length and r is the scatterer-to-receiver distance. Through numerical integration we can use this formula to obtain the predicted energy envelope of the precursor wavetrain. We compute k from the

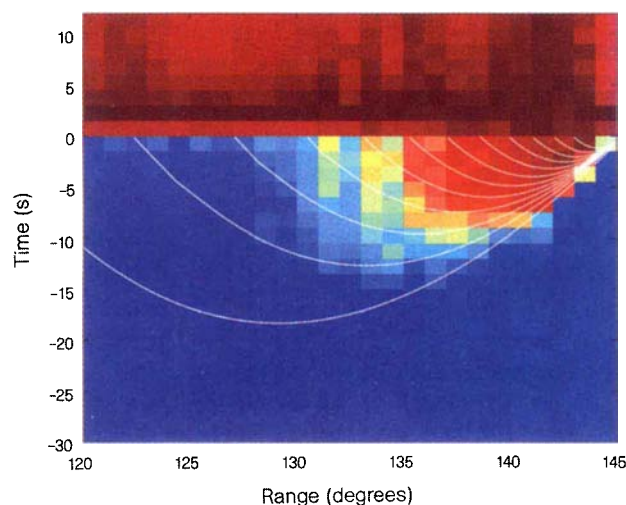


Figure 3 Average amplitude of PKP_{dif} and its precursors at ranges from 120° to 145°, resulting from a stack of 1,600 short-period seismograms. Time is relative to the onset of PKP_{dif}. Note the gradual growth in precursor amplitudes with increasing time and range. Included in this display are minimum travel-time curves for scattered arrivals calculated using the PREM velocity model³⁹. The lowest curve indicates the minimum travel time of energy single-scattered at the CMB (no energy scattered at the CMB could arrive earlier than the lowest curve). The other curves represent the minimum travel time for scattering occurring at 200-km intervals up into the mantle, with the last curve representing scattering 2,200 km above the CMB. These curves show that at short ranges and early times only scatterers close to the CMB can cause precursors, while at longer ranges and later times an increasing fraction of the mantle can contribute to the precursor wavetrain. For example, no precursors recorded at ranges less than 127° can result from scattering more than 400 km above the CMB, whereas at 143° the last part of the precursor wavetrain before the PKP_{dif} arrival could contain contributions from scattering up to 2,200 km above the CMB.

dominant 1.3-Hz frequency of our data and the P-wave velocity at the scattering depth. The $1/r^2$ factor is replaced by the geometric spreading factors and CMB transmission coefficients for the source–scatterer and scatterer–receiver ray paths. We use simple ray theory to compute these spreading factors; more sophisticated analysis has shown that this is adequate for this problem²³.

We include the contributions from both PKP_{ab} and PKP_{bc} scattering. To compare with our observations, we normalize our synthetic wavetrain to the predicted peak amplitude of PKP_{df} + PKP_{cd}, by raytracing through PREM and assuming an inner core Q of 360 (ref. 24). The choice of inner-core Q is significant as it affects the dependence of PKP_{df} amplitude on range. As we will see in the next section, the range dependence of the precursor-to-PKP_{df} amplitude ratio largely determines which scale length provides the best match between synthetics and data. Note that equation (1) predicts that the scattering becomes strongly forward for large-scale heterogeneity ($ka \gg 1$), and that the precursor amplitude varies directly with the velocity perturbation.

The depth extent of mantle scattering

To constrain the depth range of any mantle scattering, we divide the mantle into thin (20-km) layers, from the CMB to the crust, and

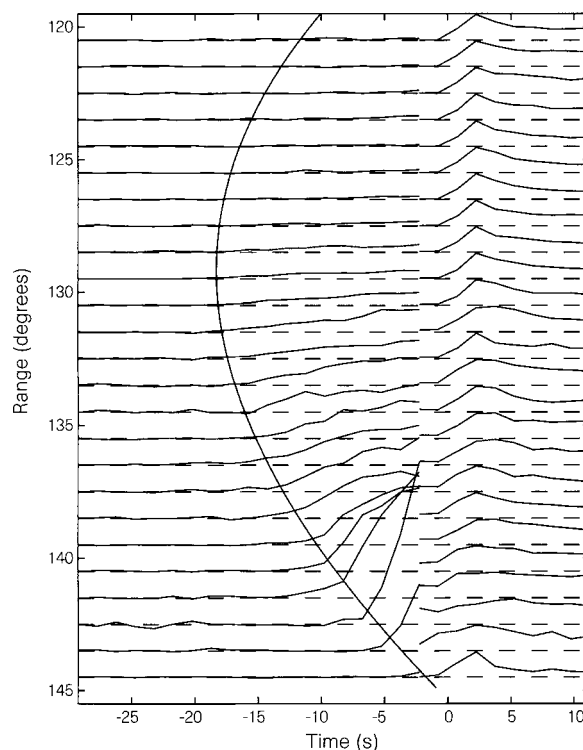


Figure 4 Average precursor amplitudes as a function of time and source–receiver range as obtained from a waveform stack of 1,600 seismograms. Times are relative to the onset of PKP_{dif}. Each stacked trace is normalized to the peak PKP_{dif} amplitude (see at about 2 s); amplitudes at negative times in this plot are multiplied by 10 to enhance the visibility of the precursors. The theoretical onset time from the PREM velocity model for single-scattering at the CMB is shown as the curved line. Note the increase in precursor levels with range and the gradual increase in precursor energy on each trace up to the time of the PKP_{dif} arrival. It is this time dependence in the precursor amplitudes that favours models of scattering distributed throughout the mantle. At the longer ranges the stacked precursor envelope tends to begin rising somewhat earlier than the PREM predicted onset time. Although some of this is due to the finiteness of our binning scheme, it may also reflect slight inaccuracies in the PREM model. Predicted PKP_{dif} precursor onset times vary by up to several seconds between different reference velocity models (for example, PREM; iasp91³⁹; sp6⁴⁰). We account for these uncertainties by permitting small time shifts in our modelling procedure (see text).

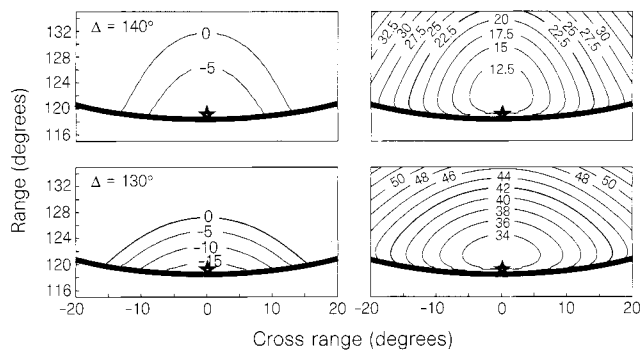


Figure 5 The two contoured plots on the left show residual travel times for energy single-scattered from PKP_{bc} to P at the CMB when compared with PKP_{gr} arrival times. At a range, Δ , of 130° the first arriving energy is expected 18.45 s before PKP_{gr} and is scattered from a point adjacent to the B caustic in the diametral plane of the source and receiver (star at a cross-range of 0° in lower left figure). The curves labelled "0" indicate the locations from which single-scattered energy will arrive at the receiver at the same time as PKP_{gr}. On the conjugate side of the travel path there are equivalent contours which describe the isotimes associated with near-source scattering. These curves are the traces, on the CMB, of three-dimensional isotime surfaces which extend into the mantle. The two plots on the right show the PKP_{bc}-to-P scattering angles at the CMB for the same grid of scatterers and receiver locations. The greatest scattering angles are associated with out-of-plane scattering. Scattering angles are much larger at shorter ranges and generally increase with time at a given range.

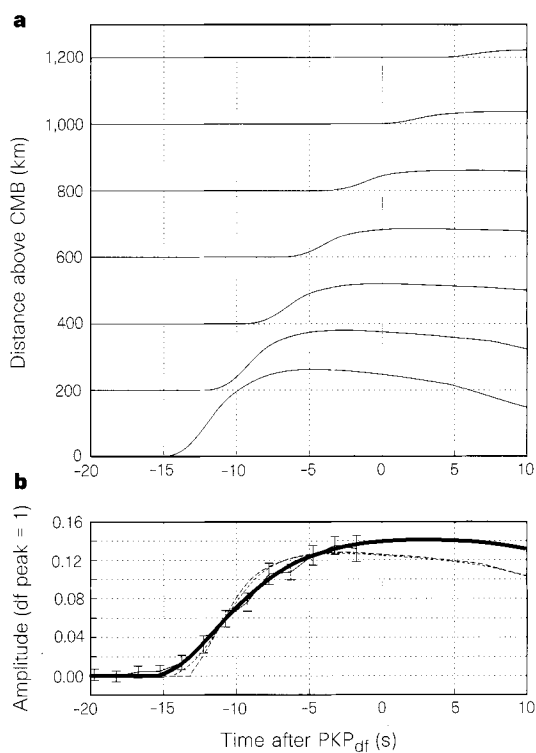


Figure 6 a. Scattering energy kernels for random mantle velocity perturbations at seven depths ranging from the CMB up to 1,200 km above it. The kernels were calculated assuming an exponential autocorrelation function with an 8-km scale and a source–receiver range of 135.5°, and have been convolved with an empirical source–time function derived from the PKP_{gr} arrivals in the stack (Fig. 4). The energy in the source–time function is concentrated in the earliest 6 s but spans 20 s to include the coda following the main arrival (probably caused by near source and receiver crustal resonance). The square root of a sum of these energy kernels can be used to model the data amplitudes displayed in Figs 3 and 4. **b.** Plot showing data fitting. An unweighted sum of all kernels yields the solid curve. This curve tracks the data much more closely than the curves that result from models that allow scattering only in the lower 200 km of the mantle (short dashes) and in the lowermost 20 km of the mantle (long dashes). The latter two models predict abrupt onsets and a levelling-off in amplitude before the PKP_{gr} arrival. One standard error bars are shown for the data curve; these are estimated through a random resampling of the seismograms.

calculate the response of each individual layer to the incident wavefield. The statistics of the heterogeneity are invariant within each layer and are allowed to change only in strength between layers. In Fig. 6 we show these responses, or kernels, for seven of these layers ranging in depth from the CMB to 1,200 km above. These were calculated assuming an exponential autocorrelation function with an 8 km scale-length and a source–receiver range of 135.5° . We find that the largest and earliest contributions will come from the deepest layers and that, at this range, the precursors must originate from a volume within 1,000 km of the CMB. Scattering that occurs at higher levels will arrive after PKP_{df} and be obscured by the coda of this phase. The amplitude decay with time present in each kernel is largely due to an increase in the scattering angle.

By summing the energy of different combinations of these kernels we can explore the effect of scattering within different mantle depth intervals. Figure 6 shows the fit obtained for whole-mantle scattering (solid line), compared to scattering restricted to the lowermost 200 km (short dashed line) and lowermost 20 km (long dashed line). We have adjusted the velocity perturbation and allowed a small time shift (to account for the uncertainties in the reference velocity model) to achieve the best fit. In this case we find that the whole-mantle scattering model provides the best match to the data.

The two models that allow heterogeneities only at the CMB (acting as a proxy for topography) or in a 200-km-thick D'' layer provide a poorer fit to the data, in that they predict a rapid increase in amplitude that tapers off just before the arrival of PKP_{df}. Only the full-mantle model allows significant contributions to the precursor wavetrain to arrive late and thus produces a curve which rises gradually and continuously until well after the arrival of PKP_{df}.

This behaviour is also seen in our observations at other ranges (Fig. 4). Precursor amplitudes increase gradually with time until the PKP_{df} arrival and do not exhibit a sharp onset that would be diagnostic of scattering restricted to the lowermost mantle. In Fig. 7 we plot our theoretical fits to 11 independent precursor stacks at ranges between 130° and 140°, showing that consistently the best fit is provided by the model that assumes uniform scattering intensity throughout the mantle. Only this model matches the emergent onset of the precursors and the continual rise in energy with time present in all bins. Considering the 143 data points in Fig. 7 within 20 s of the PKP_{df} onset; a χ^2 test returns values of 172.3, 286.0 and 373.0 for the whole-mantle, D'' and CMB models, respectively, indicating that the whole-mantle model fits the data within the standard errors (significant at the 95% confidence level), while the other models fit the observations much less well. The r.m.s. velocity

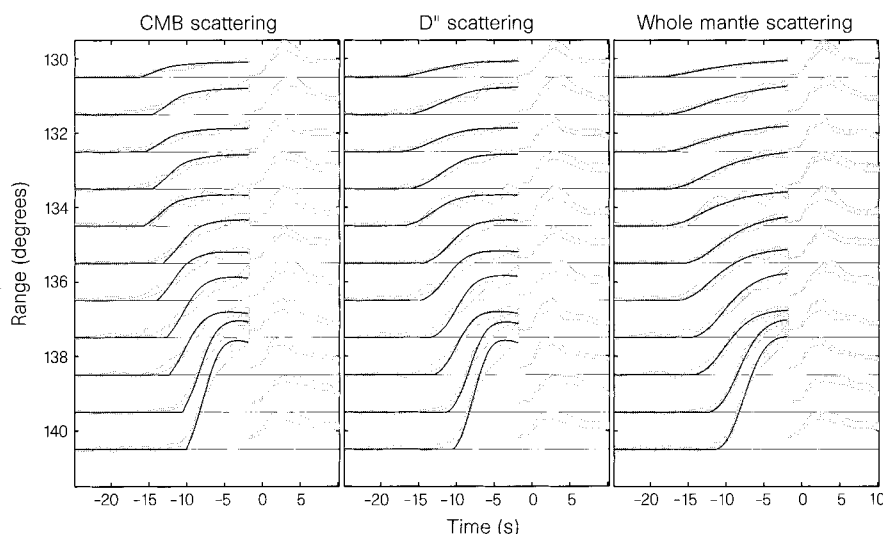


Figure 7 Observed average PKP_{diff} precursor amplitudes at ranges between 130° and 141° compared to theoretical predictions of three different models of mantle heterogeneity. The data and their standard errors are indicated with the grey shaded regions (compare with Figs 3 and 4). The left-hand panel shows the best fit to the data achieved with a model in which scattering is restricted to the core-mantle boundary. The centre panel indicates the fit when scattering is permitted within the lowermost 200 km of the mantle (that is, within the D'' region), and the

right-hand panel shows the fit obtained from a model in which the scattering strength is uniform throughout the mantle. Only the whole-mantle model provides an adequate overall fit to the data; the other models do not match the emergent onset of the precursors and their steady increase in amplitude until the main PKP_{diff} arrival. This conclusion is supported by χ^2 tests of the significance of the theoretical fits—only the whole-mantle fits the data within their standard errors.

contrast required to bring the full-mantle curves into agreement with the stack averages just over 1%. In several of the bins (for example 134.5°) the stack contains small-scale structure that is not reproduced by any of the models. In addition, several of the bins require larger velocity contrasts (the scatter was from 0.8% to 2.0%, with the amplitudes greatest in the bin spanning 131° to 132°). Because of clustering of events, some bins are dominated by certain stations or source regions; thus these anomalies may be caused by lateral variations in scattering strength in the mantle. Although this analysis shows that the CMB and D'' scattering models (similar to those proposed in the past to explain PKP precursor data) are inconsistent with the observations, we cannot exclude the possibility that more complicated models, perhaps involving multiple scattering, might exist that could explain the data without requiring mid-mantle scattering.

We prefer the 8-km scale length as it gives the best overall fit to the observations. Different scale lengths (4–16 km) can produce acceptable fits to the data, but require that the velocity contrast be strongly dependent on range. This finding is dependent on several assumptions, most notably the Q of the inner core (changing Q would favour a different scale length). However, any uncertainty in the scale length does not affect our conclusion regarding whole-mantle scattering, as the CMB and D'' scattering predictions (shown in Fig. 7) do not produce adequate fits to the data at any scale length. In addition to the exponential autocorrelation results shown here, we have also experimented with a gaussian autocorrelation model²², and find that it gives very similar results, again favouring whole-mantle scattering. Although other possible models of random heterogeneity could be considered, all will generally predict a decrease in scattered amplitude with increasing scattering angle at the low angles involved. As scattering angle increases with time at each range (Fig. 5), shallow heterogeneity (and the late precursors it provides) is required to reproduce the emergent precursor amplitudes and continuous energy increase seen in the observations.

The principal advantage of our analysis over earlier studies of PKP precursors is that we are able to resolve the time dependence of the average precursor amplitude. This time dependence is critical to resolving the question of the depth extent of the scatterers. Previous

analyses of these precursors focused mainly on modelling the range- and scattering-angle dependence of precursor amplitudes, and thus could not unambiguously determine the depth of the scattering. Our ability to image scattering in the upper mantle (above about 1,500 km depth) is limited as only a small part of the image constrains this part of the model. However, we see no evidence that scattering decreases in the upper mantle and prefer the uniform scattering model as the simplest explanation for our observations. Although more complicated models could certainly be devised, involving depth dependence in the strength and scale length of the scatterers, it is apparent that no laterally uniform model in which scattering is restricted to the lowermost mantle will fit the observations. There is no evidence from PKP precursors for an increase in small-scale heterogeneity near the CMB.

Although the focus of this Article has been PKP scattering, widespread mantle heterogeneities will cause scattering of other seismic phases and this provides a possible way to test our proposed model. Preliminary calculations of PP scattering indicate that our whole-mantle model predicts scattered arrivals between diffracted P and PP, but at such small amplitudes that these PP precursors would not be observable in our data stacks.

An important question still remains. How much topography on the CMB do the data allow? To place an upper bound on CMB r.m.s. topography we calculate the scattering response of an Earth model that contains no mantle volume heterogeneities except for a thin layer of heterogeneities at the CMB (serving as a proxy for topography). By scaling the strength of heterogeneities in this thin layer we can find the maximum velocity contrast that is within the data confidence limits and then convert this contrast to the equivalent r.m.s. topography¹⁵. Considering all bins from 120.5° to 142.5° and a horizontal scale length of 8 km, we find that the smallest upper bound (yielded by the bin spanning 124° to 125°) is 300 m. This result represents a conservative limit on CMB r.m.s. topography as any contribution from mantle scattering would reduce the allowed topography. The implication of a fairly smooth CMB at small scales is consistent with observations of high-frequency core-reflected phases^{25,26}. However, there remains evidence for at least some short-wavelength CMB topography from

PKKP precursor studies^{27–29}. Our results also do not exclude the possibility of significant CMB topography at long wavelengths.

Geodynamical implications

By stacking a large set of global data, we have produced an image of PKP precursors that shows consistent patterns in the range- and time-dependence of the precursor amplitudes. The average precursor energy levels are emergent and increase gradually in amplitude up to the arrival of the main PKP_{df} phase. This behaviour is consistent with uniform scattering through the mantle, as models in which the scattering is restricted to the D'' region predict much more impulsive precursor wavetrains. The observations do not favour a significant contribution to the precursors from CMB topography, and limit allowed relief to less than 300 m at short scales. Other than marking the base of the mantle, there is nothing in the data to suggest that there is anything special about the CMB region in regard to generating PKP precursors.

This contrasts with seismic results at longer wavelengths, where increasing evidence suggests that the D'' region is more heterogeneous than the mid-mantle^{30–34}. Our results indicate mantle velocity perturbations close to 1% at correlation lengths of 8 km, close to the dominant wavelength of our data. Extensions of these results to other scale lengths require assumptions about the heterogeneity power spectrum that are difficult to verify due to the limited bandwidth in which the PKP precursors can be observed. At long wavelengths (>500 km), both global and regional seismic tomography studies indicate a sharp fall off in mantle heterogeneity with

angular order^{31,35} suggesting that the mantle becomes more homogeneous at shorter scale lengths. Our results indicate that this fall off cannot be extrapolated to predict mantle heterogeneity at very short scale lengths.

If high frequency seismic waves are scattered throughout the mantle, then at least some of the high-frequency coda following major seismic arrivals at teleseismic distances is caused by mantle heterogeneity. However, this scattered energy may be difficult to isolate from the strong near-surface contributors to seismic coda. Our results support models containing small-scale compositional heterogeneity in the mantle^{17–19}. Subducting oceanic lithosphere entrained by mantle convection provides a possible origin for this heterogeneity¹⁷; numerical convection experiments indicate that such heterogeneity could survive for billions of years due to imperfect mixing in the flow field^{19,36}. Small-scale thermal anomalies seem a less probable source for scattering as they would quickly dissipate at short wavelengths and would have to be widespread and active on short timescales to explain our observations.

Although our focus here has been on global average precursor amplitudes, it is clear that scattering intensity varies on a regional scale. These variations are often seen on individual seismograms, with certain source or receiver regions consistently associated with more, or less, energy than the global average. Thus it is likely that some parts of the mantle scatter more strongly than the average values. We are currently attempting to characterize these intensity variations by back-propagating, or migrating, the recorded energy into the mantle³⁷. □

Received 29 April 1996; accepted 24 March 1997.

- Gutenberg, B. & Richter, C. F. On seismic waves: I. *Gerlands Beitr. z. Geophysik*, **43**, 56–133 (1934).
- Bullen, K. E. & Burke-Gaffney, T. N. Diffracted seismic waves near the PKP caustic. *Geophys. J. Int.* **1**, 9–17 (1958).
- Gutenberg, B. The boundary of the Earth's inner core. *Eos* **38**, 750–753 (1957).
- Bolt, B. A. Gutenberg's early PKP observations. *Nature* **196**, 122–124 (1962).
- Sacks, I. S. & Saa, G. The structure of the transition zone between the inner core and the outer core. *Yb. Carnegie Inst. Wash.* **69**, 419–426 (1969).
- Haddon, R. A. W. Corrugations on the CMB or transition layers between inner and outer cores? *Eos* **53**, 600 (1972).
- Cleary, J. R. & Haddon, R. A. W. Seismic wave scattering near the core–mantle boundary: a new interpretation of precursors to PKP. *Nature* **240**, 549–551 (1972).
- Doornbos, D. J. & Husebye, E. S. Array analysis of PKP phases and their precursors. *Phys. Earth Planet. Inter.* **5**, 387–399 (1972).
- King, D. W., Haddon, R. A. W. & Cleary, J. R. Array analysis of precursors to PKIKP, in the distance range 129° to 142°. *Geophys. J. R. Astron. Soc.* **37**, 157–173 (1974).
- Husebye, E. S., King, D. W. & Haddon, R. A. W. Precursors to PKIKP and seismic wave scattering near the mantle–core boundary. *J. Geophys. Res.* **81**, 1870–1882 (1976).
- Doornbos, D. J. Characteristics of lower mantle inhomogeneities from scattered waves. *Geophys. J. R. Astron. Soc.* **49**, 541–542 (1976).
- Doornbos, D. J. & Vlaar, N. J. Regions of seismic wave scattering in the Earth's mantle and precursors to PKP. *Nature* **243**, 58–61 (1973).
- Haddon, R. A. W. & Cleary, J. R. Evidence for scattering of seismic PKP waves near the mantle–core boundary. *Phys. Earth Planet. Inter.* **8**, 211–234 (1974).
- Doornbos, D. J. On seismic-wave scattering by a rough core–mantle boundary. *Geophys. J. R. Astron. Soc.* **53**, 643–662 (1978).
- Bataille, K. & Flatté, S. M. Inhomogeneities near the core–mantle boundary inferred from short-period scattered PKP waves recorded at the global digital seismograph network. *J. Geophys. Res.* **93**, 15057–15064 (1988).
- Bataille, K., Wu, R. S. & Flatté, S. M. Inhomogeneities near the core–mantle boundary evidenced from scattered waves: a review. *Pure Appl. Geophys.* **132**, 151–173 (1990).
- Davies, G. F. Geophysical and isotopic constraints on mantle convection: an interim synthesis. *J. Geophys. Res.* **89**, 6017–6040 (1984).
- Davies, G. F. Mantle plumes, mantle stirring and hotspot chemistry. *Earth Planet. Sci. Lett.* **99**, 94–109 (1990).
- Gurnis, M. & Davies, G. F. Mixing in numerical models of mantle convection incorporating plate tectonics. *J. Geophys. Res.* **91**, 6375–6395 (1986).
- Efron, B. & Tibshirani, R. Statistical data analysis in the computer age. *Science* **253**, 390–395 (1991).
- Chernov, L. A. *Wave Propagation in a Random Medium* (trans. Silverman, R. A.) (McGraw-Hill, New York, 1960).
- Wu, R. S. & Aki, K. Elastic wave scattering by a random medium and the small-scale inhomogeneities in the lithosphere. *J. Geophys. Res.* **90**, 10261–10273 (1985).
- Cornier, V. Time domain modeling of PKIKP precursors for constraints on the heterogeneity in the lowermost mantle. *Geophys. J. Int.* **212**, 725–736 (1995).
- Bhattacharyya, J., Shearer, P. M. & Masters, G. Inner core attenuation from short-period PKP(BC) versus PKP(DF) waveforms. *Geophys. J. Int.* **114**, 1–11 (1993).
- Menke, W. Few 5–20 km corrugations on the core–mantle boundary. *Geophys. Res. Lett.* **13**, 1501–1504 (1986).
- Vidale, J. E. & Benz, H. M. A sharp and flat section of the core–mantle boundary. *Nature* **359**, 627–629 (1992).
- Chang, A. C. & Cleary, J. R. Precursors to PKKP. *Bull. Seismol. Soc. Am.* **68**, 1059–1079 (1978).
- Chang, A. C. & Cleary, J. R. Scattered PKKP: further evidence for scattering at a rough core–mantle boundary. *Phys. Earth Planet. Inter.* **24**, 15–29 (1981).
- Doornbos, D. J. The effect of a rough core–mantle boundary on PKKP. *Phys. Earth Planet. Inter.* **21**, 351–358 (1980).
- Woodward, R. L. & Masters, T. G. Lower-mantle structure from ScS-S differential travel times. *Nature* **352**, 231–233 (1991).
- Su, W.-J. & Dziewonski, A. M. Predominance of long-wavelength heterogeneity in the mantle. *Nature* **352**, 121–126 (1991).
- Masters, T. G., Johnson, S., Laske, G. & Bolton, H. A shear-velocity model of the mantle. *Phil. Trans. R. Soc. Lond. A* **354**, 1–27 (1996).
- Lay, T. Structure of the core–mantle transition zone: a chemical and thermal boundary layer. *Eos* **70**, 49, 54–55, 58–59 (1989).
- Loper, D. E. & Lay, T. The core–mantle boundary region. *J. Geophys. Res.* **100**, 6397–6420 (1995).
- Passier, M. L. & Snieder, R. K. On the presence of intermediate-scale heterogeneity in the mantle. *Geophys. J. Int.* **123**, 817–837 (1995).
- Christensen, U. R. & Hofmann, A. W. Segregation of subducted oceanic crust in the convecting mantle. *J. Geophys. Res.* **99**, 19867–19884 (1994).
- Hedlin, M. A. H., Shearer, P. M. & Earle, P. Imaging CMB scatterers through migration of GSN PKP_{df} precursor recordings. *Eos* **76**, F402 (1995).
- Dziewonski, A. & Anderson, D. L. Preliminary reference earth model. *Phys. Earth Planet. Inter.* **25**, 297–356 (1981).
- Kennett, B. L. N. (ed.) *IASPEI 1991 Seismological Tables* (Res. Sch. Earth Sci., Aust. Nat. Univ., Canberra, 1991).
- Morelli, A. & Dziewonski, A. M. Body wave traveltimes and a spherically symmetric P- and S-wave velocity model. *Geophys. J. Int.* **112**, 178–194 (1993).

Acknowledgements. Data were collected by the IRIS Global Seismographic Network and obtained from the IRIS FARM. This work was supported by the US National Science Foundation.

Correspondence should be addressed to M.A.H.H. (e-mail: hedlin@ucsd.edu).

Altered cell differentiation and proliferation in mice lacking p57^{KIP2} indicates a role in Beckwith–Wiedemann syndrome

Pumin Zhang^{*†}, Nanette J. Liégeois^{†‡}, Calvin Wong^{*§}, Milton Finegold^{||}, Harry Hou[‡], Janet C. Thompson^{*§}, Adam Silverman[‡], J. Wade Harper^{*}, Ronald A. DePinho[‡] & Stephen J. Elledge^{*§†}

^{*} Verna & Marrs McLean Department of Biochemistry, [§] Howard Hughes Medical Institute, ^{||} Department of Pathology, and [†] Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA

[‡] Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, New York, New York 10461, USA

[†] These authors contributed equally to this work.

Mice lacking the imprinted Cdk inhibitor p57^{KIP2} have altered cell proliferation and differentiation, leading to abdominal muscle defects; cleft palate; endochondral bone ossification defects with incomplete differentiation of hypertrophic chondrocytes; renal medullary dysplasia; adrenal cortical hyperplasia and cytomegaly; and lens cell hyperproliferation and apoptosis. Many of these phenotypes are also seen in patients with Beckwith–Wiedemann syndrome, a pleiotropic hereditary disorder characterized by overgrowth and predisposition to cancer, suggesting that loss of p57^{KIP2} expression may play a role in the condition.

Cell proliferation in the embryo is controlled by an intricate network of extracellular and intracellular signalling pathways that process growth regulatory signals and integrate them into the basic cell-cycle regulatory machinery through control of cyclin-dependent kinases (CDKs). CDKs are positively regulated by cyclins and negatively regulated by CDK inhibitory proteins, CKIs^{1,2}. Cyclins D and E function in the G1 phase of the cell cycle and phosphorylate and inactivate the tumour-suppressor retinoblastoma (Rb) and the related proteins p107 and p130, which are negative regulators of the E2F transcription factors that facilitate cell-cycle entry³. There are two classes of CKIs in mammals, the p21^{CIP1} and p16^{INK4} families. Members of the family p16^{INK4} bind and inhibit only Cdk4 and Cdk6 kinases. Mice lacking the tumour-suppressor p16^{INK4} show no gross developmental abnormalities but have high rates of spontaneous tumorigenesis⁴. In contrast, p21^{CIP1}, p27^{KIP1} and p57^{KIP2} CKIs inhibit all G1/S phase CDKs^{1,2}. p21^{CIP1} is transcriptionally regulated by the p53 tumour-suppressor in response to DNA damage, and is required for the G1 DNA-damage checkpoint¹². Although p21^{CIP1} expression during development correlates with terminally differentiating tissues⁵, mice lacking p21^{CIP1} develop normally⁶. p27^{KIP1}-deficient mice are grossly normal developmentally but display several phenotypes that seem to be linked to cell proliferation^{7–9}. They are larger than wild-type mice, have intermediate-lobe pituitary hyperplasia or adenomas, and females are infertile.

The most structurally complex member of the p21^{CIP1} family of CKIs is p57^{KIP2} (refs 10, 11), which resides at 11p15.5 (ref. 10), a site of frequent loss of heterozygosity in several human cancers including those of the breast, bladder, lung, ovary, kidney and testicle. Several types of childhood tumours display a specific loss of maternal 11p15 alleles, suggesting the involvement of genomic imprinting¹². Rearrangements at 11p15 are seen in Beckwith–Wiedemann syndrome (BWS), which is characterized by numerous growth abnormalities, including macroglossia (enlarged tongue), gigantism, enlarged adrenal glands, ear creases, visceromegaly (enlarged organs), omphalocele (umbilical hernia), kidney abnormalities, advanced ageing and thickening of long bones, and a 1,000-fold increase in the risk of childhood tumours¹³, including

Wilms' tumour, adrenocortical carcinoma, and hepatoblastoma. BWS occurs with an incidence of 1 in 13,700 births, and 15% of these are familial with maternal carriers, suggesting a role for genomic imprinting¹⁴. BWS has highly variable penetrance in which patients usually display only a subset of all phenotypes. BWS has a complex pattern of inheritance including uniparental disomy (paternal), paternal trisomy of chromosome 11p15.5, paternal duplication of the 11p15.5 region, translocations involving maternal 11p15.5 and karyotypically normal transmission.

The protein p57^{KIP2} is encoded by a maternally expressed, imprinted gene in both humans¹⁵ and mice¹⁶, and one cluster of BWS translocations¹⁷ is within 80 kilobases of the p57^{KIP2} gene. A recent study reported that two of nine patients with BWS examined¹⁸ were found to be heterozygous for mutations in the p57^{KIP2} gene. However, there was no analysis of how frequently these alterations were observed in an asymptomatic population. Furthermore, only one patient had an apparent null allele that was shown to be maternally inherited. Because these mutations were found at such low frequency, it is possible that other genes might also be altered in these two patients. To assess directly the role of p57^{KIP2} in development, cancer and BWS, we have generated a mouse lacking the p57^{KIP2} gene. These mice have a variety of developmental defects consistent with a causative role for p57^{KIP2} in BWS, and indicate a role for p57^{KIP2} in control of cell proliferation and differentiation.

Targeted disruption of mouse p57^{KIP2} gene

A targeting construct that removed exons 1 and 2 (Fig. 1a) (87% of the p57^{KIP2} coding region) was introduced into AB2.1 embryonic stem cells. G418/gancyclovir-resistant cells were screened for homologous recombination by Southern blot analysis (Fig. 1b). Homologous recombinant cells were injected into blastocysts from C57BL/6 mice, and male chimaeras were mated to C57BL/6 females. Germline transmission was confirmed by Southern blotting (Fig. 1b). F₁ heterozygous animals were back-crossed to C57BL/6 mice to maintain the disrupted allele. To avoid ambiguity, we indicate the parental origin of alleles in heterozygotes by a superscript m for maternal or p for paternal.

The disrupted allele is a null as demonstrated by the absence of $p57^{KIP2}$ mRNA (Fig. 1e) and protein (Fig. 1c,e) in $p57^{-/-}$ or $p57^{+/-m}$ animals. $p57^{KIP2}$ is detected in the tectum of brain, kidney, adrenal gland, muscle, lung and cartilage in wild-type embryos. Levels of $p21^{CIP1}$ and $p27^{KIP1}$ were unchanged in tissues from $p57^{KIP2}$ mutants, except in muscle where a slight increase in $p27^{KIP1}$ was detected (Fig. 1c). The $p57^{+/-p}$ animals had wild-type expression of $p57^{KIP2}$.

Generation of $p57^{KIP2}$ mutant mice

Of 32 offspring from $p57^{+/-p}$ female to $p57^{+/+}$ male matings, no $p57^{+/-}$ animals were present when genotyped at two weeks of age (Table 1). Of 82 offspring from $p57^{+/-p}$ intercrosses, 55% were

$p57^{+/+}$ and 45% were $p57^{+/-}$ (Table 1); no $p57^{-/-}$ animals survived to two weeks of age. Mendelian inheritance predicts a 1:2 ratio for $p57^{+/+}$ to $p57^{+/-}$ animals, indicating that half of the $p57^{+/-}$ animals died before genotyping. We observed dead or dying new-born mice in several litters, and these were found to be mutant. We concluded that $p57^{KIP2}$ is required for postnatal survival in a hybrid C57BL/6-129Sv background. Although lethal in this background, we recently discovered that crossing to the outbred CD1 strain allows some $p57^{+/-m}$ animals to survive well beyond day one.

When evaluated between embryonic day (E)18.5 and E20, genotypes were detected at expected Mendelian frequencies (Table 1). However, 10% of mutant embryos were dead, staged from E13 to E16, consistent with the fact that maternal inheritance of the $p57^{KIP2}$

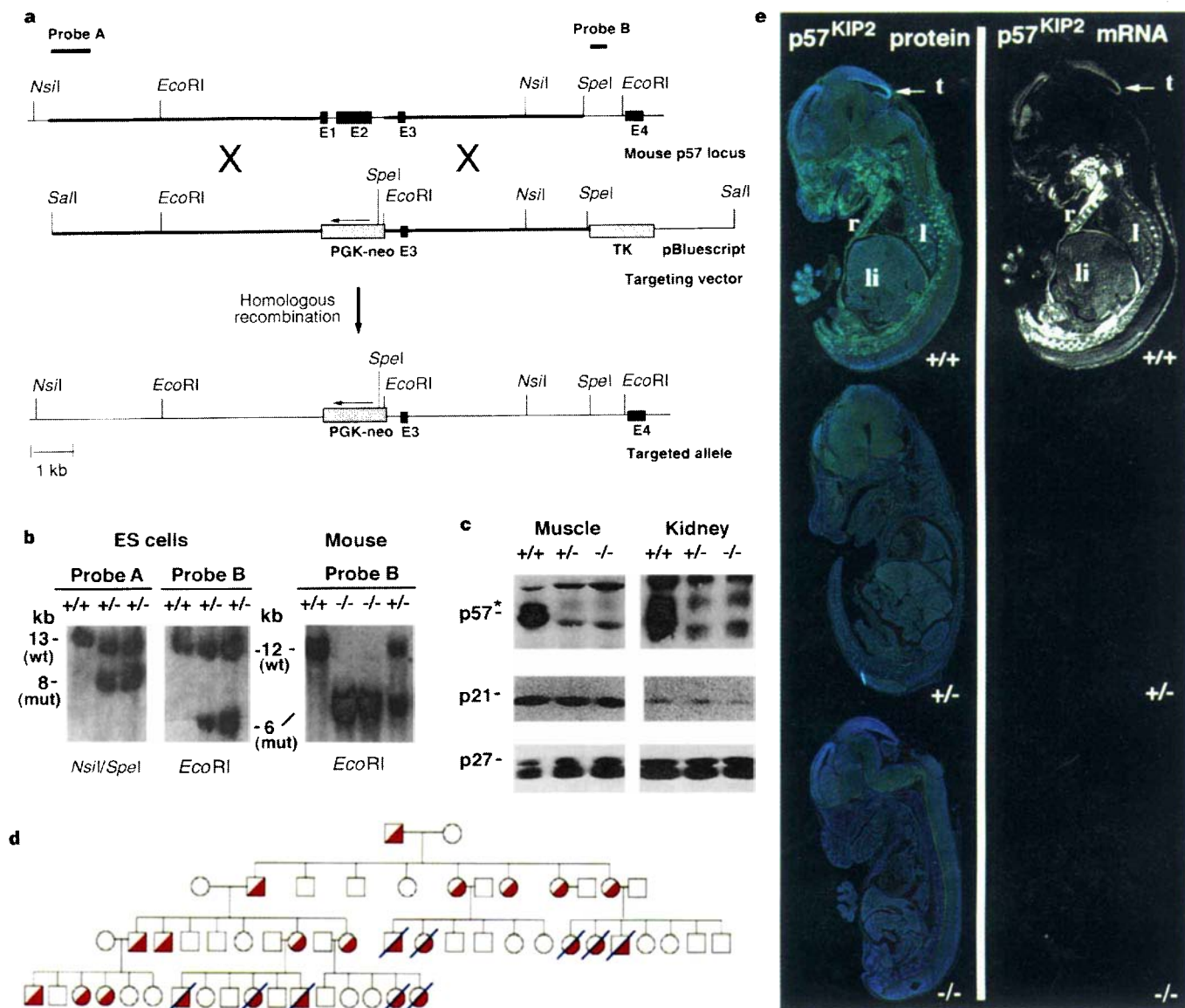


Figure 1 Targeted disruption of $p57^{KIP2}$. **a**, $p57^{KIP2}$ disruption strategy. Probes (A and B) for Southern analysis are indicated. **b**, Southern blot analysis of DNA from wild-type and mutant ES clones and embryos. **c**, Western blot analysis of $p57^{KIP2}$, $p21^{CIP1}$ and $p27^{KIP1}$ proteins in muscle and kidney. The asterisk indicates a form of $p57^{KIP2}$ resulting from phosphorylation or alternative splicing. **d**, A pedigree analysis: squares, males; circles, females. Heterozygotes are represented by

half-filled symbols. Animals displaying a mutant phenotype are indicated by diagonal lines through symbols. **e**, *In situ* hybridization and immunofluorescent analysis of sagittal sections derived from E15.5 $p57^{+/+}$, $p57^{+/-m}$ and $p57^{-/-}$ embryos. A combined image of $p57^{KIP2}$ protein (green) and nuclei stained with Hoechst dye (blue) is shown; l, lung; li, liver; r, rib; t, tectum.

null allele is lethal (Fig. 1d). The $p57^{-/-}$ animals were phenotypically indistinguishable from affected $p57^{+/-m}$ heterozygotes.

Mice lacking $p57^{KIP2}$ have omphalocele

Mutant embryos showed umbilical abnormalities as early as E16.5. A herniated abdomen was noticeable in all mutants (Fig. 2A, d,e) together with malrotation of the intestines resulting in placement of the jejunum and ileum in front of the liver. Occasionally, the small intestines were found outside the abdominal cavity (omphalocele), a range of phenotypes characteristic of BWS in humans (Fig. 2A, b). Most dead or dying neonates had a slit in their abdomen where the umbilicus is normally positioned, and portions of their visceral organs were missing, presumably devoured by their mother in the process of removing the placenta and yolk sac (Fig. 2A, c).

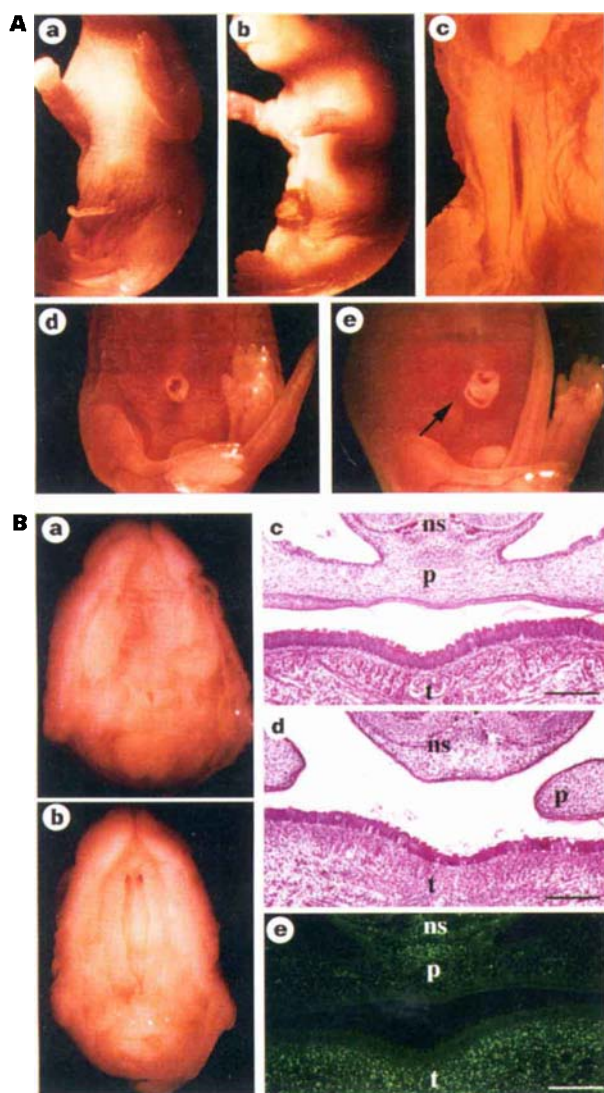


Figure 2 Omphalocele, umbilical hernia and cleft palate in $p57^{KIP2}$ mutant mice. **A**, Intestines are found outside the abdominal cavity in some $p57^{KIP2}$ mutant newborns (**b**) but not wild-type mice (**a**). Umbilical sacs were removed in **b** to expose intestines. Typically, dead neonates (mutants) had a slit on their abdomen with missing viscera (**c**). Umbilical hernia is observed in mutants (**e**, arrow) but not in wild-type littermates (**d**). **a**, **b**, E20; **c**, P0; **d**, **e**, E18. **B**, Jaws were removed from P0 wild-type (**a**) and mutant (**b**) mice to allow palate viewing. Haematoxylin and eosin-stained coronal sections show a fused palate in E20 wild-type (**c**) mice and cleft palate in mutant (**d**). $p57^{KIP2}$ expression (green) in the palate mesenchyme and tongue muscle of an E20 embryo (**e**); ns, nasal septum; p, palate; t, tongue. Scale bars, 200 μm.

The small intestines develop outside the body until E15.5 (ref. 19), when then they enter the body cavity. How this occurs is unknown. It is unclear how omphalocele occurs in BWS, but the prevailing assumption is that visceral overgrowth limits abdominal space. In $p57^{KIP2}$ mutant embryos, visceral overgrowth was not evident. $p57^{KIP2}$ is highly expressed in intestinal mesenteries that form connections between the intestine folds and the abdomen and may participate in intestine re-entry. However, no obvious histological malformation was detected in mutant mesenteries.

Body wall muscle dysplasia in $p57^{KIP2}$ mutants

E18 mutants are 10% shorter than weight-matched $p57^{+/+}$ embryos (Fig. 3A, a). However, skeletons are nearly identical lengths (Fig. 3A, b), indicating a different aetiology for the body-length anomaly. Because proper musculature is required for skeletal stature, we investigated muscle organization. Histological examination of mutant embryos revealed defects in the position of body wall muscles (Fig. 3B, b,d). At the umbilicus level, the muscle did not reach as far towards the midline of the abdomen in mutants as in wild-type mice, leaving large areas of the abdominal wall uncovered by muscle. To distinguish whether this was a cause or a consequence of herniation, we examined E14.5 embryos in which a physiological omphalocele is normally present. Muscle in the mutants also failed to reach the midline (Fig. 3B, a,c), strongly suggesting that the abdominal wall defect is responsible for the subsequent umbilical herniation.

Overall muscle differentiation was not affected in $p57$ -mutant mice, as indicated by the normal muscle-fibre organization and peripheral location of nuclei. Surprisingly only 50% of muscle nuclei express $p57^{KIP2}$ (Fig. 3B, e,f). This is consistent with the hypothesis that there are two muscle lineages, MyoD and Myf5. Perhaps $p57^{KIP2}$, which is also expressed in E11.5 somites (data not shown), is expressed in only one lineage, and is required for muscle-cell migration.

Neonatal lethality in $p57^{KIP2}$ mutant mice

Cleft palate (secondary palate) and difficulty in breathing were noticeable in all mutant neonates (Fig. 2B, compare a and b). Milk was found in their lungs, and air in their stomach and intestines, causing inflation and stretching. No histological abnormalities were found in mutant diaphragm, lung, bronchi or trachea. The severity of cleft palate was variable but could compromise breathing by allowing an accumulation of liquid in the nasopharynx that is subsequently brought into the lungs by inhalation.

Closure of the secondary palate is a complex process²⁰. Vertically growing palate shelves elevate by E13 and grow together, fusing by E16 to form the secondary palate¹⁹. Analysis of E20 mutant embryos demonstrated that the palate failed to fuse but maintained normal organization (Fig. 2B, c,d). $p57^{KIP2}$ is expressed in mesenchymal cells of the palate and muscle cells of the tongue, but not nasal or oral epithelia (Fig. 2B, e). Failure of palate closure could result from defects in mesenchymal cell migration, response to induction signals, cell proliferation, or increased apoptosis caused by inappropriate proliferation. The tongue of mutant mice is not enlarged, and is therefore not interfering in palate closure.

Renal medullary dysplasia in $p57^{KIP2}$ mutants

A major histopathological finding in BWS is non-cystic medullary dysplasia and enlargement of the kidney²¹. BWS medullary pyramids are often poorly formed and exhibit abundant connective tissue stroma with widely separated renal tubules²². In $p57^{KIP2}$ mutants the size and organization of the kidney was normal (Fig. 4a–h), but the inner medullary pyramid was significantly smaller than normal. Normally at E18.5 and E20 the maturing glomeruli are juxtamedullary and nephrons have developed long loops of Henle reaching deep into the inner medullary region (Fig. 4g). However, mutants have fewer renal tubules (loops of Henle and collecting

Table 1 Genotype frequencies of offspring

* Male p57 ^{+/+} × female p57 ^{+/−} p			
p57 genotype	+/+	+/−	
Number	32	0	
Observed (%)	100	0	
Expected (%)	50	50	
* Male p57 ^{+/−} - p × female p57 ^{+/−} - p			
p57 genotype	+/+	+/−	− / −
Number	45	37	0
Observed (%)	55	45	0
Expected (%)	25	50	25
† Male p57 ^{+/+} × female p57 ^{+/−} - p			
p57 genotype	+/+	+/−	
Number	24	27	
Observed (%)	47	53	
Expected (%)	50	50	
† Male p57 ^{+/−} - p × female p57 ^{+/−} - p			
p57 genotype	+/+	+/−	− / −
Number	16	27	14
Observed (%)	28	47	25
Expected (%)	25	50	25

* Genotyped at 2 weeks of age.

† Genotyped between E18.5 and E20.

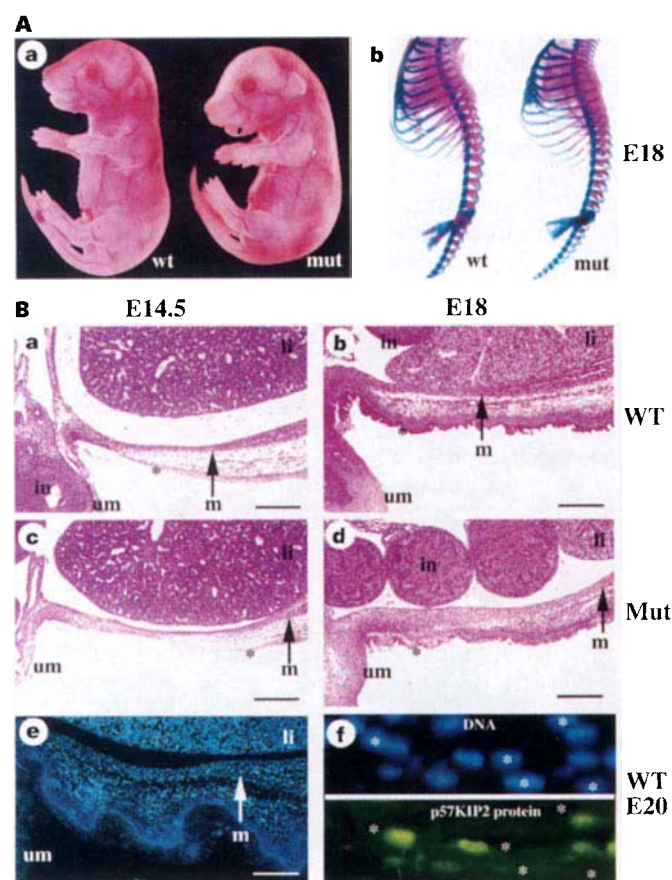


Figure 3 Body-wall dysplasia in p57^{KIP2} mutants. **A**, E18 p57^{+/+}(wt) and p57^{+/−} m(mut) embryos (**a**) and skeletons (**b**) prepared from the same embryos. Mutants have shorter bodies but normal skeletal length. **B**, Transverse embryo sections were stained with haematoxylin and eosin. The distance between umbilicus and the tip of rectus abdominis muscle (arrows) is greater in mutants than in wild-type littermates in E14.5 (**a** and **c**) and E18 (**b** and **d**) embryos. Skin development is also delayed in mutants (asterisks). p57^{KIP2} protein was detected in the rectus abdominis muscle, connective tissue, skin, and liver (**e**), and is present in half of the nuclei in muscle (**f**); in, intestine; li, liver; m, muscle; um, umbilical cord. Scale bars, 200 μ m.

ducts) and more stromal cells in the inner medulla (Fig. 4d,h), a phenotype very similar to that of BWS patients²¹. No defects in primary nephrogenesis were observed, indicating that ureter bud branching and mesenchymal induction occurred normally in the mutant kidney.

To determine whether medullary dysplasia is due to improper development or to regression of a properly formed kidney, a developmental analysis was performed. At E16 the renal pelvis forms through fusion of initial ureter branches. A normal renal pelvis was observed in the E16.5 mutant kidney (Fig. A, B). At this stage the inner medulla normally begins to form, but in the mutant this was totally absent. At E18.5 the mutant inner medulla has also failed to develop properly (Fig. 4C, D). This indicates that the medullary dysplasia is due to arrested or improper development.

p57^{KIP2} is expressed in podocytes of maturing glomeruli and in interstitial stromal cells between renal tubules (Fig. 4i) but not the tubules themselves. This expression pattern argues against an intrinsic defect in the renal tubule, and underscores the importance of tissue–tissue interactions between epithelia of renal tubules and surrounding mesenchymal cells during the development of the loops of Henle, as has been demonstrated previously for ureteric branching and the mesenchymal to epithelial transition.

Bone development and collagen X expression

Skeletal staining with alcian blue (cartilage) and alizarin red (bone) revealed abnormalities in mutant skeletons. At E14.5 the two sternal bands in mutants were widely separated, whereas those in wild-type animals had already fused (Fig. 5A, a,b). By E19, wild-type sternbrae were fully ossified, whereas ossification had just begun in the mutant (Fig. 5A, c,d). Two separate ossification centres in each sternbrae were clearly seen in the mutant, indicating imperfect fusion of sternal bands (Fig. 5A, d). Smaller ossification centres were observed in mutants at all stages of development in forelimbs, vertebra and supraoccipital bone, as shown in Fig. 5A, e–l. Although mutant limbs are shorter, they are thicker than in the wild type (compare Fig. 5A, g and h).

Vertebrate long bones are formed through endochondral ossification, which involves formation of a cartilage framework that is converted to bone by replacement. Within this framework, cells are organized into distinct zones: the epiphyseal centre zone contains resting chondrocytes that act as stem cells for the adjacent proliferative zone, where chondrocytes proliferate and form columns, and the hypertrophic zone, containing dying chondrocytes that are in the process of ossification. Histological examination of mutant long-bone sections showed a slight disorganization of the columnar alignment of differentiating chondrocytes (Fig. 5B, a,b). Furthermore, the mutant hypertrophic zone is slightly thinner than in the wild type and contains smaller cells, suggesting impairment of chondrocyte differentiation.

p57^{KIP2} is expressed at moderate levels in resting chondrocytes, low levels in the proliferative zone, and very high levels in the hypertrophic zone (Fig. 5B, c). We examined cell-cycle withdrawal in the mutant hypertrophic zone to determine whether cell division in this zone might account for resistance to ossification. A higher level of BrdU incorporation was observed in the resting (2.2-fold) and proliferative (1.6-fold) chondrocytes of E15 mutant animals, but at later times, such as postnatal day (P)0, only a small increase in BrdU labelling was observed in resting (20%) and proliferative (14%) chondrocytes. At E18.5 there is a 10% greater cell density in these two zones, which may explain the bone thickening in mutant animals (Fig. 5A, e–h). However, no BrdU labelling was observed in either the mutant or wild-type hypertrophic zones at any stage.

Collagen X is expressed in hypertrophic chondrocytes and has been implicated in proper bone development²³. Expression of collagen X was significantly reduced in the mutant hypertrophic zone (Fig. 5B, d,e). Thus p57^{KIP2} is required for expression of collagen X, and perhaps other genes that facilitate the ossification

of chondrocytes. Because no cell-cycle entry was observed in mutant hypertrophic chondrocytes, the failure to express collagen X suggests that $p57^{KIP2}$ may play a direct role in processes of differentiation.

Adrenal cortex hyperplasia and cytomegaly

The adrenal gland is among the most consistently enlarged organs in BWS patients and shows extensive cytomegaly. $p57^{+/-m}$ mutants show a significant enlargement of the adrenal gland (a 25% to 100% increase in volume; Fig. 6a). Although mutant adrenals were normal histologically, there was a threefold increase in the frequency of cytomegaly (Fig. 6c,d). Expression of $p57^{KIP2}$ is restricted to fetal adrenal cortex (Fig. 6b), indicating a role in controlling cell proliferation.

Cycling and apoptosis in lens after loss of $p57^{KIP2}$

The $p57^{KIP2}$ protein is present in high levels in postmitotic lens fibre cells, with low-level, sporadic expression in the anterior epithelial layer (Fig. 7a,b). $p57^{KIP2}$ is induced in the equatorial zone (Fig. 7a, arrows) where epithelial cells are withdrawing from the cell cycle and initiating terminal differentiation²⁴, suggesting a role for $p57^{KIP2}$ in promoting one or both of these processes. Analysis of $p57^{-/-}$ and $p57^{+/-m}$ lenses revealed grossly normal lens structure at E13.5 with minor vacuolization, but a more pronounced vacuolization in E15.5 and older lenses (Fig. 7e). Late-stage differentiation markers such as γ -crystallin and membrane-intrinsic protein-26 were unaffected (data not shown). However, $p57^{KIP2}$ loss causes inappropriate S-phase entry in lens fibre cells (Fig. 7f,g) and an increase in apoptotic nuclei (Fig. 7i,j), possibly contributing to the vacuolated appearance. Another phenotype of $p57^{-/-}$ lenses was a 10-fold increase in apoptosis in the anterior epithelial compartment (Fig. 7j). Epithelial cells that accumulate in the centralmost position of the anterior epithelial layer are normally eliminated by a p53-independent apoptotic mechanism to maintain the single-cell layer anteriorly²⁵. Because these cells express $p57^{KIP2}$, the increase in apoptosis could result from an accelerated rate of proliferation and execution of a normal mechanism designed to eliminate excess cells in the central zone of the anterior lens.

Discussion

$p57^{KIP2}$ acts as a regulator of cell proliferation in the adrenal gland, the lens epithelia and certain chondrocytes. The partial dependency on $p57^{KIP2}$ for reducing cell proliferation reveals the redundant mechanisms used to limit tissue growth. A similar situation is observed in cell culture where agents that induce cell-cycle arrest immediately increase levels of certain CKIs and subsequently reduce the levels of the cyclins and CDKs. While undergoing the process of reducing CDK activity during differentiation, the absence of CKIs may allow additional cell cycles to occur before CDK activity is sufficiently reduced to block cell-cycle entry.

Loss of Rb is associated with increases in proliferation, impaired expression of differentiation markers, and inappropriate apoptosis in lens fibre cells²⁵. $p57^{-/-}$ lenses show less cell proliferation and apoptosis than $Rb^{-/-}$ lenses, thus $p57^{KIP2}$ is likely to play a partly redundant role upstream of Rb. However, the increase in apoptosis in the anterior epithelial compartment is greater in $p57^{-/-}$ than in $Rb^{-/-}$ lenses. Thus, even in the same cell type, the relationship between $p57^{KIP2}$ and Rb (and possibly other cell-cycle regulators) changes with respect to differentiation state. Furthermore, the ossification defect in $p57^{-/-}$ mice is similar to that observed in $p107^{-/-}p130^{-/-}$ double mutants²⁶, and it is likely that $p57^{KIP2}$ is also regulating these proteins to control proliferation in chondrocytes.

Several tissues in $p57^{-/-}$ mutants showed developmental defects not obviously linked directly to increased cell proliferation, such as defects in kidney development, differentiation of hypertrophic chondrocytes, muscle organization, and formation of the secondary

palate. These defects may indicate roles in differentiation distinct from the biochemical role of $p57^{KIP2}$ as a CKI. In the case of the ossification defects, the hypertrophic chondrocytes exit the cell cycle properly but were disorganized and failed to express differentiation markers such as collagen X. Non-CDK binding regions of the $p57^{KIP2}$ protein, such as the repeat regions or the conserved QT domains, may play cell cycle-independent roles in differentiation.

Using the $p57^{KIP2}$ -defective mouse allows us to assess unambiguously the role of $p57^{KIP2}$ in BWS. Its BWS-related phenotypes support a causal role for a mutation found previously¹⁷. Both patients with $p57^{KIP2}$ mutations displayed omphalocele, macroglossia, gigantism and earlobe grooves. $p57^{KIP2}$ -deficient mice also show omphalocele but not gigantism, macroglossia, visceromegaly and hypoglycaemia (data not shown). However, the expression pattern of $p57^{KIP2}$ is consistent with a possible role in both gigantism

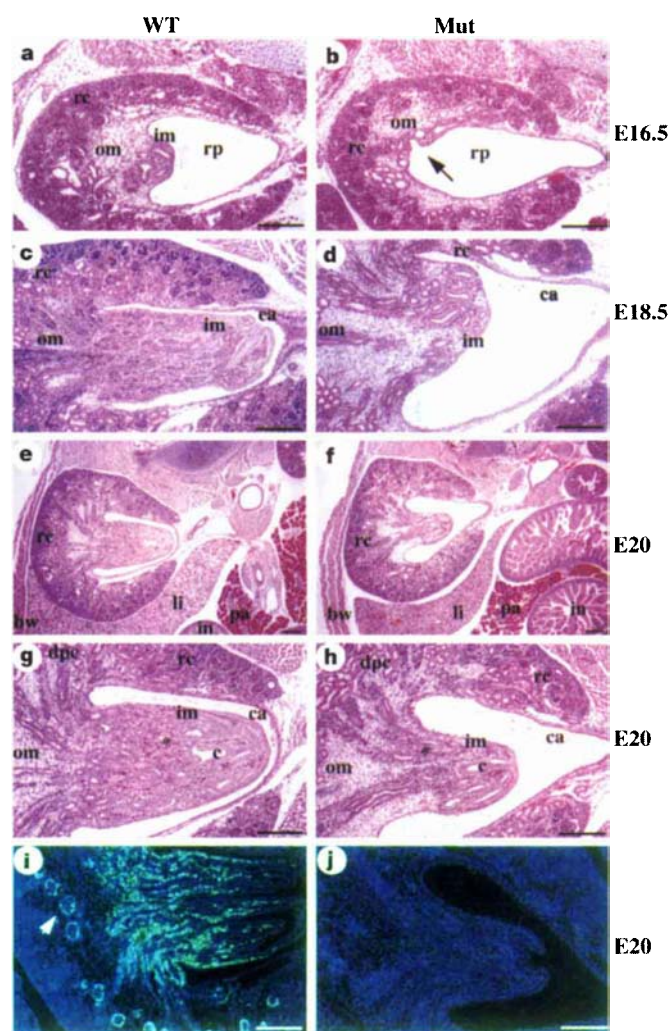


Figure 4 Kidney medullary dysplasia in $p57^{KIP2}$ mutants. **a, b**, At E16.5, the inner medulla is absent in the $p57^{-/-}$ mutant kidney (arrow). **c, d**, At E18.5, a delayed elongation of the inner medulla is evident in the mutant kidney. **e, f**, Low magnification shows relatively normal organization of the mutant kidney at E20. **g, h**, A higher magnification of **e** and **f**. The mutant inner medulla is significantly smaller than in the wild type, containing fewer loops of Henle (asterisks) and collecting ducts with more stromal cells in between. **(i, j)** $p57^{KIP2}$ protein is expressed in mesenchymal cells between renal tubules of the inner medulla and in podocytes of glomeruli (arrow), but not in mutants (**j**); bw, body wall; c, collecting ducts; dpc, distal and proximal convoluted tubules; im, inner medulla; in, intestine; li, liver; om, outer medulla; pa, pancreas; ca, renal calyx; rc, renal cortex; rp, renal pelvis. Scale bars, 200 μ m.

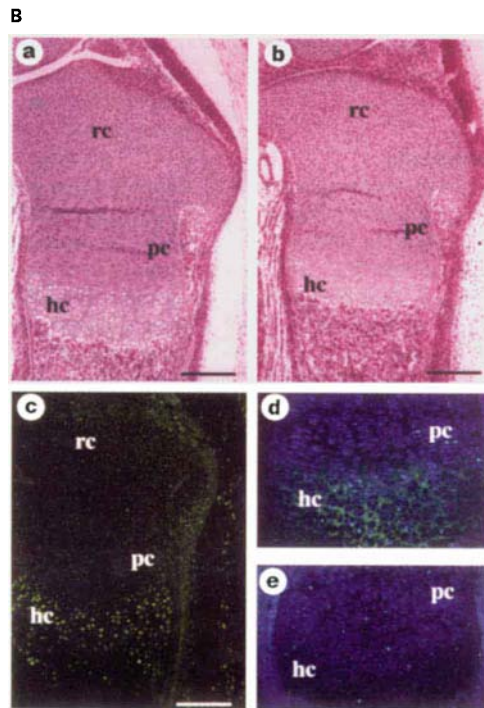
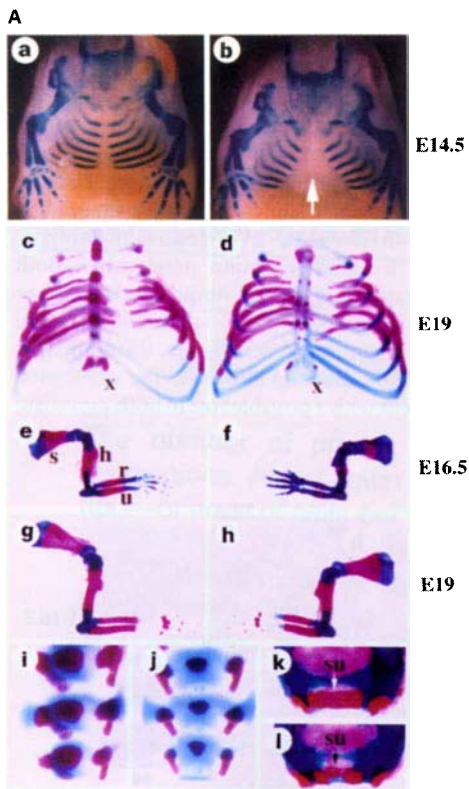


Figure 5 Impaired endochondral ossification in $p57^{KIP2}$ mutants. **A**, Alcian blue (cartilage) and alizarin red (bone) staining of E14.5 skeletons showing fused sternum rudiments in wild-type (**a**) and an unfused sternum in $p57^{KIP2-/-}$ mice (**b**, arrow). Sternebrae are fully ossified in E19 wild-type (**c**) but not mutant (**d**) embryos. Note the separate sternum ossification centres and enlargement of the xiphoid process in **d**. **e–h**, Forelimbs of E16.5 (**e**, **f**) and E19 (**g**, **h**) of wild-type (**e**, **g**) and mutant (**f**, **h**) embryos. Reduced ossification is observed in mutant E19 vertebrae (**j**) and P0 supraoccipital bone (**l**) relative to wild-type (**i**, **k**). **B**, Longitudinal sections through the tibia of E18.5 wild-type (**a**) and mutant (**b**) embryos, stained with haematoxylin and eosin. $p57^{KIP2}$ expression was visualized by immunofluorescence (**c**). Collagen X expression was detected in wild-type (**d**) but not in mutant (**e**) hypertrophic chondrocytes; h, humerus; hc, hypertrophic chondrocytes; pc, proliferating chondrocytes; r, radius; rc, reserving chondrocytes; s, scapula; su, supraoccipital bone; u, ulna; x, xiphoid process. Scale bars, 200 μ m.

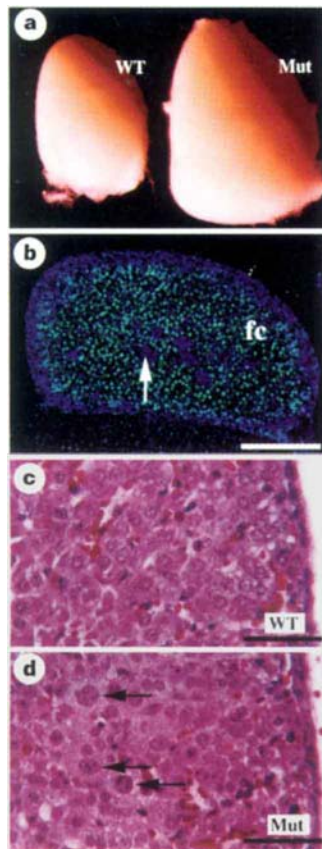


Figure 6 Hyperplasia of the adrenal gland in $p57^{KIP2}$ mutants. **a**, The E19 adrenal gland is enlarged in $p57^{KIP2-/-}$ mice. **b**, $p57^{KIP2}$ protein is expressed in the fetal cortex but not the medulla (arrow) in an E20 adrenal gland. **c**, **d**, Haematoxylin and eosin-stained sections of E20 wild-type (**c**) and $p57^{KIP2-/-}$ (**d**) adrenal glands. Arrows indicate cytomegaly; fc, fetal cortex; m, medulla. Scale bars: **b**, 200 μ m; **c**, **d**, 50 μ m.

and macroglossia; the mouse tongue has extremely high levels of $p57^{KIP2}$ during development, as does the human placenta, which controls nutrient flow to embryos^{10,11}.

It is unknown which of the internal manifestations of BWS these two patients exhibit. However, the $p57^{KIP2-/-}$ mouse displays several common internal BWS phenotypes including enlargement of the adrenal cortex, adrenal cytomegaly, and renal medullary dysplasia, which are now strongly predicted to be present in the two affected individuals. Phenotypes present in the $p57^{KIP2-/-}$ mouse not frequently considered to be associated with BWS are cleft palate, endochondral ossification defects, and ocular lens impairment. It should be noted that, unlike the mouse, human $p57^{KIP2}$ imprinting is incomplete with 5% residual expression from the paternal allele¹⁵. Therefore, although mice lacking the maternal allele are null mutants, the equivalent humans are hypomorphs. This may affect the relative penetrance of phenotypes between species. Nevertheless, increased frequency of cleft palate^{27,28}, skeletal anomalies²⁹, and cataract formation²⁹ have been reported in conjunction with BWS. A very high frequency of palate defects including cleft and soft palate have been observed in 6 of 10 BWS patients in a study from Japan²⁸. The high frequency of palate defects in that study relative to others may reflect the differences in frequencies of modifier genes in distinct populations. In response to this study it was learned that the BWS patient from a previous study¹⁸ who had a maternally inherited null allele in $p57^{KIP2}$ also had a cleft palate (I. Hatada, personal communication). This provides further compelling support for the involvement of $p57^{KIP2}$ in BWS. It has also been observed that the long bones in BWS patients often show widened metaphyses and a thickened bony cortex²¹. This is consistent with our observations of the thickened bones in $p57^{KIP2-/-}$ deficient mice. Thus the analysis of the $p57^{KIP2}$ -deficient mouse has not only provided proof of the role of $p57^{KIP2}$ in BWS, but has also extended the range of phenotypes expected on further analysis of BWS patients with $p57^{KIP2}$ mutations. Furthermore, it has provided an explanation for the physiological basis of several phenotypes observed in BWS patients, including skeletal abnormalities (chondrocyte proliferation and differentiation), omphalocele (abdominal

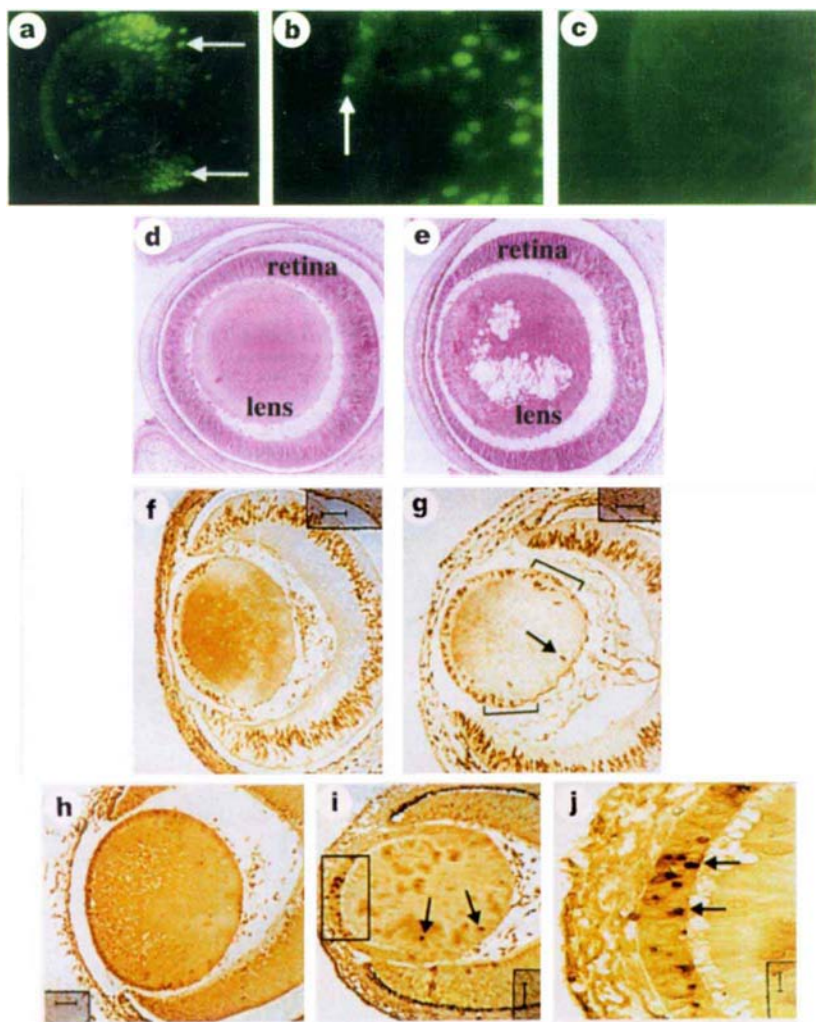


Figure 7 Loss of p57^{KIP2} causes increased proliferation and apoptosis in the lens. **a–c**, Immunofluorescent analysis shows strong p57^{KIP2} expression in all nuclei of the post-mitotic lens fibre cell compartment (arrows) (**a**), occasional nuclear and cytoplasmic staining in anterior epithelial cells (arrow) (**b**), and no expression in p57^{-/-} lenses (**c**). **d, e**, Histological presentation of haematoxylin and eosin-stained E15.5 lens sections (sagittal) derived from p57^{+/+} (**d**) or p57^{-/-} (**e**) embryos. **f, g**, BrdU incorporation assays on E13.5 lens sections derived from p57^{+/+} (**f**) or p57^{-/-} (**g**) embryos. Brackets show inappropriate S-phase entry in normally postmitotic lens fibre cells of the equatorial region. The arrow indicates S-phase entry in central lens fibre cells. **h–j**, TUNEL assays on E13.5 lens sections derived from p57^{+/+} (**h**), p57^{-/-} (**i**), p57^{-/-} (**j**) embryos (higher magnification of the boxed region in **i** showing abundant apoptotic nuclei in the anterior epithelial layer). Scale bars: **f–i**, 10 μm; **j**, 4 μm.

muscle developmental defects) and adrenomegaly (absence of a cell proliferation inhibitor).

Because p57^{KIP2} is central to BWS, it is likely that most or all BWS forms affect p57^{KIP2} function in some way. Loss of p57^{KIP2} could account for BWS involving alterations at the maternal 11p15.5 locus. However, it is more difficult to imagine how p57^{KIP2} is directly involved in the form of BWS caused by duplication of the paternal 11p15.5 region. It is possible that duplicating the imprinted paternal p57^{KIP2} allele could affect the maternal allele's expression, but this seems unlikely. An alternative explanation is the linked-antagonist model, which proposes that mutations in p57^{KIP2} or overexpression of a p57^{KIP2} antagonist cause BWS. The chromosome specificity of the paternal duplication could be accounted for by exclusive paternal expression of the antagonist, or by the fact that, owing to tight linkage, the duplicated region would also carry p57^{KIP2}, and so duplications including the maternal p57^{KIP2} would not produce BWS. Under these conditions, duplication of the paternal 11p15 region would have an effect similar to mutating the p57^{KIP2} gene, increasing a dosage-dependent antagonist. It is possible that IGF2 is the linked antagonist that acts in the same pathway as p57^{KIP2}, but in opposition. Increased IGF2 expression leads to some phenotypes associated with BWS in mice³⁰ and humans³¹. The linked-antagonist model predicts that each mutation (or increased dosage) would have some common effects (for example, omphalocele, macroglossia and gigantism) and some unique effects (for example, earlobe creases, viceromegaly and renal medullary dysplasia), and there would be significant variability in the range of different phenotypes, depending on which genes were mutant or over-

produced. Paternal uniparental disomy would cause the most severe phenotype because all p57^{KIP2} expression would be lost and increased dosage of the antagonist(s) would occur. The linked-antagonist model accounts for the highly variable penetrance associated with this syndrome.

Because we now have an understanding of the molecular defects that cause BWS, there should be a concerted effort to correlate phenotype with genotype of BWS patients to establish whether groups of phenotypes are linked and correlate with a particular genotype. This analysis should help establish whether BWS is a single disease or a collection of similar diseases that operate through an overlapping pathway, and will definitively establish whether p57^{KIP2} is a tumour suppressor. □

Methods

Targeting vector construction. The p57^{KIP2} targeting vector was constructed using plasmids containing *PGK-neo* and *pMC1-HSV-tk*. *PGK-neo* on a *XhoI-HindIII* fragment was cloned into p57Sal1d3 cleaved with *HindIII* and partly with *XhoI* to make p57KO-2A, containing a 7-kb fragment of p57^{KIP2} 5'-genomic DNA. A 5.6-kb *SpeI-EcoRV* fragment of p57^{KIP2} 3'-genomic DNA was ligated into *XbaI-ClaI* (blunt)-digested pMC1TK generating p57KO-2B with *tk* gene 3' to p57^{KIP2} genomic DNA. To generate the targeting vector p57KO-2, the *SalI/EcoRV* fragment of p57KO-2A containing p57^{KIP2} 5'-genomic DNA and *neo* was ligated into *SalI-ClaI* (blunt)-digested p57KO-2B. p57KO-2 introduces *EcoRI* and *SpeI* sites into p57^{KIP2} locus.

ES cell culture and germline transmission of targeted allele. Linearized p57KO-2 (25 μg) was electroporated into 1.1 × 10⁷ AB2.1 ES cells and cultured³². Homologous recombinant clones were microinjected into 3.5-d.p.c.

C57BL/6 blastocysts and a male chimaera derived from clone p57-2.2.12B transmitted the targeted allele.

Gross and histological analysis. Embryos were fixed in 10% formalin or Bouin's. For histological analysis, fixed samples were dehydrated using ascending concentrations of ethanol, cleared in xylene and embedded in paraffin wax. Embedded samples were sectioned at 5 μ m.

Immunohistochemistry and *in situ* hybridization. Immunostaining for collagen X was performed as described²³. For cell proliferation assays, pregnant mice were injected with BrdU (0.1 mg per g body weight) 2 h before caesarean section, and positive cells were identified with an anti-BrdU monoclonal antibody (Dako). *In situ* hybridization was performed as described³. Apoptosis and BrdU incorporation assays were performed as described²⁵.

Protein analysis. Tissues from E18 embryos were lysed in NP-40 buffer and cleared by centrifugation. Protein (30 μ g) was immunoblotted with anti-p57^{KIP2} (monoclonal antibody KP90)³³, anti-p21^{CIP1} (monoclonal antibody 65)⁶ or anti-p27^{KIP1} antibodies (provided by A. Koff) using ECL detection.

Received 16 January; accepted 3 April 1997.

- Harper, J. W. & Elledge, S. J. Cdk inhibitors in development and cancer. *Curr. Opin. Genet. Dev.* **6**, 56–64 (1996).
- Sherr, C. J. Cancer cell cycles. *Science* **274**, 1672–1677 (1996).
- Weinberg, R. A. The retinoblastoma protein and cell cycle control. *Cell* **81**, 323–330 (1995).
- Serrano, M. *et al.* Role of the INK4a locus in tumor suppression and cell mortality. *Cell* **85**, 27–37 (1996).
- Parker, S. B. *et al.* p53-Independent expression of p21^{CIP1} in muscle and other terminally differentiating cells. *Science* **267**, 1024–1027 (1995).
- Deng, C. *et al.* Mice lacking p21^{CIP1/WAF1} undergo normal development, but are defective in G1 checkpoint control. *Cell* **82**, 675–684 (1995).
- Nakayama, K. *et al.* Mice lacking p27^{KIP1} display increased body size, multiple organ hyperplasia, retinal dysplasia, and pituitary tumors. *Cell* **85**, 707–720 (1996).
- Kiyokawa, H. *et al.* Enhanced growth of mice lacking the cyclin-dependent kinase inhibitor function of p27^{KIP1}. *Cell* **85**, 721–732 (1996).
- Fero, M. L. *et al.* A syndrome of multiorgan hyperplasia with features of gigantism, tumorigenesis, and female sterility in p27^{KIP1}-deficient mice. *Cell* **85**, 733–744 (1996).
- Matsuoka, S. *et al.* p57^{KIP2}, a structurally distinct member of the p21^{CIP1} Cdk-inhibitor family, is a candidate tumor suppressor gene. *Genes Dev.* **9**, 650–662 (1995).
- Lee, M. L. *et al.* Cloning of p57^{KIP2}, a cyclin-dependent kinase inhibitor with unique domain structure and tissue distribution. *Genes Dev.* **9**, 639–649 (1995).
- Hastie, N. D. The genetics of Wilms' tumor—a case of disrupted development. *Annu. Rev. Genet.* **28**, 523–558 (1994).
- Wiedemann, H. R. Tumours and hemihypertrophy associated with Wiedemann–Beckwith syndrome. *Eur. J. Pediatr.* **141**, 129–134 (1983).

- Junien, C. Beckwith–Wiedemann syndrome, tumorigenesis and imprinting. *Curr. Opin. Biol.* **2**, 431–438 (1992).
- Matsuoka, S. *et al.* Imprinting of the gene encoding a human cyclin-dependent kinase inhibitor p57^{KIP2}, on chromosome 11p15. *Proc. Natl Acad. Sci. USA* **93**, 3026–3030 (1996).
- Hatada, I. & Mukai, T. Genomic imprinting of p57^{KIP2}, a cyclin-dependent kinase inhibitor, in mouse. *Nature Genet.* **11**, 204–206 (1995).
- Hoovers, J. *et al.* Multiple genetic loci within 11p15.5 defined by Beckwith–Wiedemann Syndrome rearrangement breakpoints and subchromosomal transferable fragments. *Proc. Natl Acad. Sci. USA* **92**, 12456–12460 (1995).
- Hatada, I. *et al.* An imprinted gene p57^{KIP2} is mutated in Beckwith–Wiedemann syndrome. *Nature Genet.* **14**, 171–173 (1996).
- Kaufman, M. H. *The Atlas of Mouse Development* (Academic, London, 1992).
- Ferguson, M. J. W. Palate development. *Development* (suppl.) **103**, 41–60 (1988).
- Beckwith, B. Macroglossia, omphalocele, adrenal cytomegaly, gigantism, and hyperplastic visceromegaly. *Birth Defects: Original Article Series 2* Vol. V No. 2 (ed. Bergsma, D.) 188–196 (The National Foundation, 1969).
- Hepinstall, R. H. in *Pathology of the Kidney* Vol. 1, 114–155 (Little, Brown, Boston, MA, 1992).
- Jacenko, O. *et al.* Spondylometaphyseal dysplasia in mice carrying a dominant negative mutation in a matrix protein specific for cartilage-to-bone transition. *Nature* **365**, 56–61 (1993).
- McAvoy, J. W. Induction of the eye lens. *Differentiation* **17**, 137–149 (1996).
- Morgenbesser, S. D. *et al.* p53-dependent apoptosis produced by Rb deficiency in the developing mouse lens. *Nature* **371**, 72–74 (1994).
- Cobrinik, D. *et al.* Shared role of the Rb-related p130 and p107 proteins in limb development. *Genes Dev.* **10**, 1633–1644 (1996).
- Elliott, M. *et al.* Clinical features and natural history of Beckwith–Wiedemann syndrome: presentation of 74 new cases. *Clin. Genet.* **46**, 168–174 (1994).
- Takato, T., Kamei, M., Kato, K. & Kitano, I. Cleft palate in the Beckwith–Wiedemann syndrome. *Ann. Plastic Surg.* **22**, 347–349 (1989).
- Sotelo-Avila, C., Gonzalez-Crussi, F. & Fowler, J. W. Complete and incomplete forms of Beckwith–Wiedemann syndrome: their oncogenic potential. *J. Pediatr.* **96**, 47–50 (1980).
- Leighton, P. A. *et al.* Disruption of imprinting caused by deletion of the H19 gene region in mice. *Nature* **375**, 34–39 (1995).
- Weksberg, R., Shem, D. R., Song, Q. L. & Squire, J. Disruption of IGF2 imprinting in Beckwith–Wiedemann Syndrome. *Nature Genet.* **5**, 143–149 (1993).
- Ramirez-Solis, R. *et al.* Hoxb-4 (Hox-2.6) mutant mice show homeotic transformation of a cervical vertebra and defects in the closure of the sternal rudiments. *Cell* **73**, 279–294 (1993).
- Dynlacht, B. D. *et al.* Purification and analysis of CIP/KIP proteins. *Methods Enzymol.* (in the press).

Acknowledgements. We thank A. Koff, R. Haronen, B. Olsen, J. Jorwitz and S. Zigler for antisera; A. Bradley for advice and help with gene disruptions; F. DeMayo, S. Beckwith, M. Edwards, S. Plon, M. Blackburn, U. Albrecht and B. Smith for discussions; F. Macaluso for processing of lens samples; and S. Tilghman, A. Beaudet, V. Schuster and A. Feinberg for discussions and critically reading the manuscript. This work was supported by DAMD breast cancer grants for S.J.E. and J.W.H., an ATP grant to S.J.E. and Baylor SPORE in Prostate Cancer. P.Z. is supported by a training grant from the NIA. R.A.D. is a recipient of the Irma T. Hirsch career scientist award and is supported by NIH grants. S.J.E. is a Pew scholar in the biomedical sciences and an investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to R.A.D. or S.J.E. (sledge@bcm.tmc.edu).

YOURS TO HAVE AND TO HOLD BUT NOT TO COPY

The publication you are reading is protected by copyright law. Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931 Fax: 0171 436 3986

The peculiar colours of the halo light in the edge-on spiral galaxy NGC5907

Richard J. Rudy*, Charles E. Woodward†, Tracy Hodge†, Stephen W. Fairfield† & David E. Harker†

* Space and Environment Technology Center, The Aerospace Corporation, PO Box 92957, Los Angeles, California 90009, USA

† Wyoming Infrared Observatory, Department of Physics and Astronomy, University of Wyoming, Laramie, Wyoming 82071-3905, USA

The presence of substantial haloes of 'dark matter' around galaxies has been inferred from their gravitational influence on the gas and stars in their disks¹, but the nature of the dark matter remains very uncertain. The recent detection of faint optical emission from the halo around the edge-on galaxy NGC5907 (refs 2–5), distributed in a manner that follows the expected distribution of the gravitational mass, provided the first direct indication that faint stars might be the repository of some dark matter. But it was not clear how much of the mass was provided by these intrinsically faint stars. Here we report near-infrared observations of the halo emission from this galaxy. Taken together with the optical data, the results produce a very peculiar spectral energy distribution, which cannot be explained by any current models of stellar populations. The best approximation is a collection of stars with near-solar abundances of heavy elements, along with many low-mass stars. Such a population would be very unexpected for a galactic halo, where the stars should be old and have rather low fractions of heavy elements, but could account for much of the gravitational mass.

Infrared searches for the massive, 'dark' haloes that surround galaxies require observations reaching four orders of magnitude fainter than the ambient sky and thermal background. For this reason, separating the halo from the sky on the same image, as is done in the optical^{2,3,5}, is not possible. Instead, the telescope is shifted rapidly between the halo and surrounding sky. This technique of measuring halo and sky with the same pixels removes the need for the precise flatfields necessary in the optical^{2,3,5}, but at the expense of half the integration time. For our observations, sky was sampled every 10 seconds at distances more than 6' from the source position and at least 4' further from the plane of NGC5907. To

minimize out-of-field changes in background occurring within the observatory dome, the region surrounding the camera window was masked and the rotation of the dome was monitored.

Sky-subtracted images were corrected for pixel-to-pixel gain differences using flatfields derived from the sky observations and then transformed to absolute fluxes by measuring standard stars⁶. Using field stars for registration, we combined all images in a given bandpass to form a single mosaic centred on the nucleus and extending to 145'' perpendicular to the plane on both the southwest and northeast sides. All detectable stars were then masked (assigned zero value) and excluded from determinations of the halo brightness.

Because source and sky were not sampled simultaneously, sky subtraction was generally imperfect, resulting in an apparent offset in the flux. Although small with respect to the sky itself, this residual can be large compared with the halo emission. For this reason, we measure the gradient of the emission across the array^{4,7}, which corresponds to the change in brightness across the galaxy. Nevertheless, to derive the absolute brightness of the halo, it is still necessary to determine the 'zero' level. For our data, zero brightness levels on both sides of the plane were set equal to the intercepts at 145'' obtained by linear fits to the data between 120'' and 145''. Infrared brightnesses (Table 1; Fig. 1) were calculated in this manner. The *J* and 'K short' (*K_s*) values 95'' from the plane, 24.66 and 23.00 respectively for the southwest side, and 24.02 and 23.25 for the northeast, are in good agreement with the average values of 24.3³ and 23.0 reported by James and Casali⁴. However, setting the zero point at 145'' ignores any underlying emission that may be present and that extends beyond this distance, so all magnitudes determined with this assumption are low values. This effect is negligible close to the plane but becomes increasingly severe as the distance approaches 145''. To allow for the real possibility of emission beyond 145'', we also considered the case where the brightness falls continuously at the rate measured in the R band³, which is roughly $(r + r_{\text{core}})^{-\alpha}$, where the core radius $r_{\text{core}} = 37.4''$ and $\alpha = 1.26$. Using this relationship together with our observations of the halo flux between 75'' and 100'', we calculated flux values at 145''. These offsets were added to the values of Table 1 to produce a second set of infrared magnitudes. Because the value 60'' northeast of the plane calculated in this manner is ~ 0.2 mag brighter than the upper limit in the K band reported by Skrutskie *et al.*⁸ (their observations, made with a single detector using standard infrared sky 'chopping' techniques, do not have the uncertain offset), the second set of magnitudes probably gives an upper bound to the brightness. Although not tabulated explicitly, these magnitudes are included in Fig. 2 and in the discussion to follow.

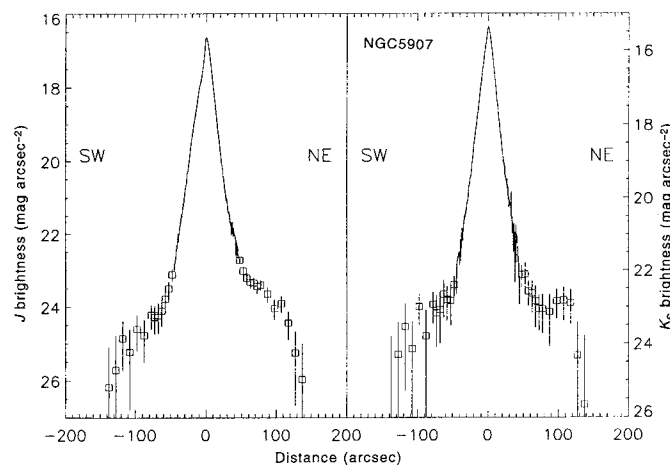


Figure 1 Brightness profiles of NGC5907 in the *J* (1.25 μm) and *K_s* (2.15 μm) filters. Data between 45'' and 70'' from the plane are grouped into 5'' bins; beyond 70'', bins 10'' wide are used. Error bars are 2σ and reflect the internal errors within the bin as well as the statistical uncertainty in the zero point. The contribution from light diffracted from the telescope, secondary mirror and its support structure was assessed from observations of the bright star μ UMa within, and outside, the camera aperture, and found to be negligible over the spatial range spanned by the observations. Detectable point sources have been excluded. The slight depression $\sim 5''$ southwest of the peak (more pronounced in the *J* data) is caused by a dust lane. Its apparent offset from the peak is due to the small inclination (3°) of the galactic plane to the line of sight. The profiles in both bandpasses are dominated by the stars of the disk. This component shows a rapid, exponential fall-off to about 45 arcsec from the plane. The scale heights for the *J* and *K_s* data, obtained from fitting a hyperbolic secant to the data, are 5.5'' and 5.0'', respectively. Beyond 45'' the faint extended emission ascribed to the halo^{3–5} begins. The asymmetries between opposite sides of the plane seen here in data from both filters are not apparent in the V and I bands⁵. James and Casali⁴ speculated that much of the emission beyond the disk might be due to the infall of companion dwarf galaxies; it is this type of asymmetric phenomenon that seems necessary to explain the bump on the northeast side of the brightness profile.

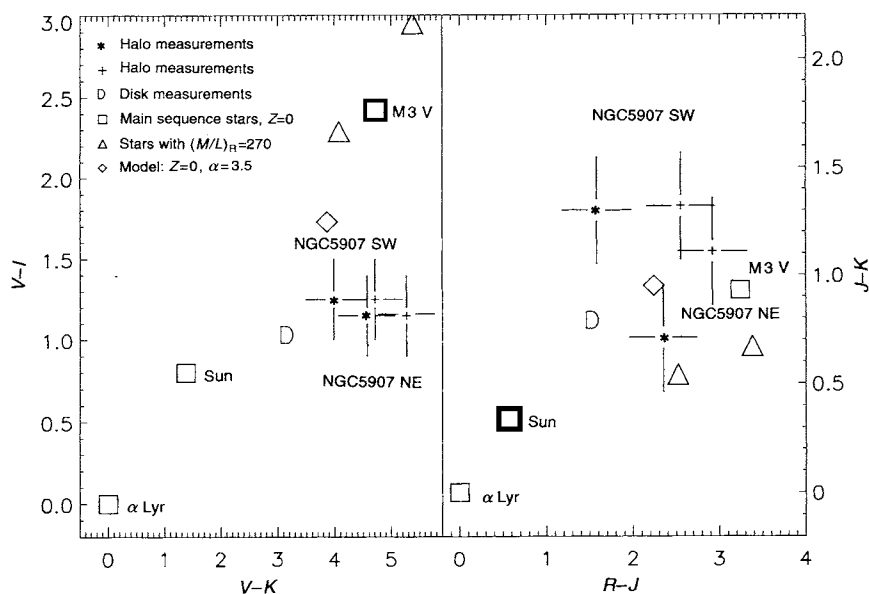


Figure 2 A comparison of observations of NGC5907 with selected stars and theoretical models. Observations of the halo, with error bars, are shown for both sides of the galaxy: * denotes values from Table 1, which assume that the halo ends at 145"; + are derived assuming that the infrared brightness declines at the same rate as the optical. Disk colours were measured 20" to 50" away from the plane, on both sides. The M3 dwarf has the minimum R-band mass-to-light ratio, $(M/L)_R = 270$, necessary to account of the rotation curve. Triangles denote stars of reduced metallicity ($Z = -0.5$, where Z is the logarithm of the ratio of the abundance of heavy elements to the solar value) which also have $(M/L)_R = 270$. Because the observations are of integrated galactic light, colours and $(M/R)_R$ values were calculated for stellar populations with various metallicities (solar, -0.5 , -1.5) and stellar mass functions: $\alpha = 1.35$ (the standard Salpeter value), 3.0

and 3.5 ($N(m) \sim m^{-(1+\alpha)}$, where $N(m)$ is the number density for objects of mass m). A single epoch of star formation was assumed; $0.08M_\odot$ was taken as the lower limit of hydrogen burning. Colours and mass-luminosity relations are from observations¹³ and from the theoretical calculations^{6,17}. For stars with $Z < 0$ and masses greater than $0.4M_\odot$, observations of globular clusters were used²⁹⁻³³. Models for zero metallicity³⁴ provided a check for the limiting case, although no colours indicative of such stars were seen (for example, $J-K = -0.3$ for an object at the limit of hydrogen burning). Colours for the giant component were derived from integrated observations of globular clusters³⁶, assuming a contribution of 70% in the R band³⁶ for $\alpha = 1.35$. The model plotted produced the best agreement with the observations.

Before considering the colours of the halo emission, which is our principal topic here, we touch briefly on the shape of the brightness profile. The shape of the profile, in particular the rate of decline in the halo region, is important because, if the distribution of matter in the halo is traced by the emission, a sharp drop in brightness could indicate too little mass far enough from the plane to produce the observed rotation curve. The rate of decline seen in Fig. 1 mimics, but does not exactly duplicate, the fall-off observed in the optical³. In addition to the asymmetry between the two sides of the plane (see Fig. 1), the brightness to the southwest declines slightly more rapidly than $r^{-1.26}$. We fitted both a disk and halo component³ to our data to determine the rate of decline. Although the fall-off at K_s ($\sim r^{-1.5}$) is close to the optical value, the J-band emission on the southwest side declines as r^{-2} (implying a mass density that falls as r^{-3}). If, however, emission does exist beyond 145", the effect of setting the flux level to zero at this distance is to steepen the apparent rate of decline. Thus the actual fall-off in flux is likely to be more gradual than we measure. Although our infrared data cannot confirm the shallow fall-off observed in the optical³, they are consistent with this result.

To gain insight into the source of the halo emission, we investigated its spectral energy distribution using the $V-I$, $V-K$, $R-J$ and $J-K$ colours. (K_s magnitudes were transformed to standard K values using profiles for both filters, knowledge of the atmospheric transmission and the $J-K_s$ colour. In all cases the changes were < 0.10 mag.) Halo colours were compiled for both sides of the galaxy by averaging the bins in Table 1 that are 60" or more from the plane. The values for the southwest and northeast sides, respectively, are: $V-I = 1.25, 1.15$; $V-K = 3.99, 4.57$; $R-J = 1.58, 2.35$; $J-K = 1.30, 0.71$. When the infrared brightness distribution is assumed to follow a $r^{-1.26}$ profile, the three colours involving

infrared magnitudes become redder: $V-K = 4.71, 5.27$; $R-J = 2.55, 2.92$; $J-K = 1.32, 1.11$. Obviously, the uncertainties for the optical-to-infrared colours are large (~ 0.7 mag), reflecting the uncertainty in the zero point and in combining data from different instruments and wavelength regimes. With regard to the latter, however, the optical-infrared colours of the disk measured from 20"-50" from the plane ($V-K = 3.11$; $R-J = 1.52$) are quite similar to values measured in other spirals^{9,10}. This is also true of the $J-K$ value of 0.80, with only the low value of $V-I = 1.03$ being anything other than typical⁹.

In marked contrast to its disk, the halo colours of NGC5907 are very peculiar. The $V-K$ and $J-K$ colours are extremely red whereas the $V-I$ colour is unusually blue. The $V-K$ colour, which is a sensitive indicator of metallicity¹¹ (higher metallicity results in larger (that is, redder) $V-K$ values), is at least one magnitude redder than the disk. It is redder than any of the values measured for the nuclei (where higher metallicity is expected) of a sample of spirals¹⁰. Values of $V-K$ greater than 4.0 are characteristic of dwarfs and giants of solar metallicity of spectral type M2 or later¹². Likewise, the $J-K$ colour measured by James and Casali⁴, and by us for the southwest component, is observed in only the latest, metal-rich stars¹³. In direct contrast to this, the $V-I$ colour is indicative of either an earlier spectral type (K0 for main sequence stars) or a lower metallicity. The value observed for the halo of NGC5907 is much bluer than a region of the bulge of the edge-on spiral NGC7814 measured at a similar distance from its plane¹⁴. The small $V-I$ colour also rules out reddening by dust grains as the cause of the large $V-K$ and $J-K$ values.

In attempting to constrain the stellar constituents of the halo, we compared the observed colours with model colours of integrated stellar distributions^{11,15}. Although none provided a good match,

Table 1 Infrared and optical magnitudes for NGC5907

Position	J	K _s	R	V	I
-145 to -120	25.64	24.48	-	-	-
-120 to -100	25.07	23.68	26.72	-	-
-100 to -80	24.59	23.25	26.31	27.58	26.36
-80 to -70	24.32	23.02	25.96	27.00	25.67
-70 to -60	24.16	22.85	25.71	26.50	25.29
-60 to -50	23.64	22.80	25.02	25.95	24.87
-50 to -40	22.80	22.04	24.15	25.04	24.09
-40 to -30	21.50	20.82	23.04	23.85	22.86
-30 to -20	20.05	19.34	21.79	22.63	21.59
-20 to -10	18.63	17.62	20.56	21.65	20.45
0					
10 to 20	18.74	17.71	20.56	20.92	19.62
20 to 30	20.40	19.40	21.79	22.41	21.10
30 to 40	21.63	20.68	23.04	23.77	22.79
40 to 50	22.48	21.84	24.15	25.06	24.15
50 to 60	23.11	22.31	25.02	26.04	25.05
60 to 70	23.32	22.74	25.71	26.68	25.60
70 to 80	23.39	22.95	25.96	27.15	25.99
80 to 100	23.76	23.13	26.31	27.67	26.52
100 to 120	24.10	22.69	26.72	28.30	27.08
120 to 145	25.38	24.88	-	-	-

Magnitudes are in units of magnitudes arcsec⁻². Distances are perpendicular to the galactic plane, in units of arcseconds. Negative values are southwest of the plane, positive are to the northeast. Infrared data, in the standard J (1.25 μ m) bandpass and in 'K short' (1.98–2.32 μ m, K_s), are from this work. R-band magnitudes are from ref. 3; because they averaged data from both sides of the galaxy, the table reports equal values on opposite sides of the plane. V and I data are from ref. 5. Infrared measurements were acquired on the nights of 3 and 5 February, 21 May, and 30 and 31 August 1995 with the Aerospace near-infrared camera (incorporating a 256 × 256 HgCdTe NICMOS3 array) and the 2.4-m reflector of the Wyoming Infrared Observatory. To derive the infrared magnitudes, a broad swath of data, corresponding to the full 110° width of the detector array, was binned. The V and I data were reported in a similar fashion⁵. Because the R-band data were originally presented in narrower slices³, the cut through the centre of the galaxy was combined with parallel cuts that were offset by 77" to yield the values here. In all cases the corrections were <0.25 mag, and in good agreement with the values computed assuming a galaxy morphology as described by ref. 28. Errors for infrared measurements: 10°–40°, 0.1 mag; 40°–60°, 0.2 mag; 60°–100°, 0.25 (J), 0.3 (K_s), 100°–145°, 0.4 (J), 0.5 (K_s).

some trends were apparent. Most notable was that the large $V - K$ value indicated a population of higher (near solar) metallicity with either a giant component that was unusually red or a very large contribution to the infrared flux from low-mass stars. To model populations of solar metallicity, incorporate the most recent colours calculated for low-mass stars^{16,17} and estimate the colours for mixes of very different populations, we also computed integrated colours (Fig. 2). None of our models produced colours identical to those measured, but the model that was most similar incorporated a population of solar metallicity, a main-sequence cutoff mass of 1.2 solar masses ($M_{\odot}$) and a very steep mass function ($\alpha = 3.5$). The resulting colours are: $V - I = 1.80$; $V - K = 4.09$; $R - I = 2.54$; $J - K = 0.93$. If the stars are older, and thus accompanied by a giant population, the colours are slightly bluer (Fig. 2). Given the measurement uncertainties, both models might describe the halo emission, except that such populations are entirely unexpected for a galactic halo. Halo formation is believed to occur early in the life of a galaxy before large amounts of metals have been manufactured. Moreover, a mass function rising as steeply as $\alpha = 3.5$ and extending down to the hydrogen burning limit is unknown in observations of either the globular clusters^{18–22} or the halo of the Milky Way^{23–26}.

Interestingly, if such a population does exist, the preponderance of low-mass stars results in a very large mass-to-light ratio (M/L). The main sequence whose colours are listed above has a M/L value of 250 in the R band, very near the minimum value of 270 necessary to explain the rotation curve³. The presence of a giant component increases the luminosity with negligible addition to the mass, but a modest extension of the mass function into the brown-dwarf regime (to $0.06M_{\odot}$) produces $(M/L)_R > 270$.

The reality of such a population, which requires that stars with masses $<0.25M_{\odot}$ be at least 20 times more numerous than for any

population observed in our Galaxy, hinges critically on the source of the infrared emission. In addition to verifying the red $V - K$ colour, future observations of the halo of NGC5907 should endeavour to determine the relative contributions to the infrared from dwarfs and giants. The H₂O and CO (refs 11, 27) absorptions would serve this purpose. If giants contribute significantly, the mass-to-light ratio is small and the dark matter must take some non-luminous form; if dwarfs generate the near-infrared emission, then the principal source of gravitational mass is stars. □

Received 8 October 1996; accepted 18 March 1997.

- Ashman, K. M. Dark matter in galaxies. *Publ. Astron. Soc. Pacif.* **106**, 1109–1138 (1992).
- Morrison, H. L., Boroson, T. A. & Harding, P. Stellar populations in edge-on galaxies from deep CCD surface photometry. *Astron. J.* **108**, 1191–1207 (1994).
- Sackett, P. D., Morrison, H. L., Harding, P. & Boroson, T. A. A faint luminous halo that may trace the dark matter around the spiral galaxy NGC5907. *Nature* **370**, 441–443 (1994).
- James, P. & Casali, M. M. Near-infrared emission from a spiral galaxy halo. *Spectrum* 14–16 (March 1995).
- Lequeux, J., Fort, B., Dantel-Fort, M., Cuillandre, J.-C. & Mellier, Y. V and I band observations of the halo of NGC5907. *Astron. Astrophys.* **312**, L1–L4 (1996).
- Elias, J. H., Frogel, J. A., Matthews, K. & Neugebauer, G. Infrared standard stars. *Astron. J.* **87**, 1029–1034 (1982).
- Casali, M. M. & James, P. A deep near-infrared search for halo emission from an edge-on spiral galaxy. *Mon. Not. R. Astron. Soc.* **274**, 265–270 (1995).
- Skrutskie, M. F., Shure, M. A. & Beckwith, S. Faint photometry of edge-on spiral galaxies: a search for massive halos. *Astrophys. J.* **299**, 303–311 (1985).
- Terndrup, D. M. et al. Infrared imaging of spiral galaxies: colors and luminosity profiles. *Astrophys. J.* **432**, 518–546 (1994).
- Frogel, J. A. The stellar content of the nuclei of late-type spiral galaxies. *Astrophys. J.* **298**, 528–543 (1985).
- Buzzoni, A. Evolutionary population synthesis in stellar systems. II. Early-type galaxies. *Astrophys. J. Suppl. Ser.* **98**, 69–101 (1995).
- Koorneef, J. Near-infrared photometry. *Astron. Astrophys.* **128**, 84–93 (1983).
- Kirkpatrick, J. D. & McCarthy, D. W. Low mass companions to nearby stars: spectral classification and its relation to the stellar/substellar break. *Astron. J.* **107**, 333–343 (1994).
- Lequeux, J., Dantel-Fort, M. & Fort, B. The bulge and the warp of NGC 7814: stars, gas, and dust. *Astron. Astrophys.* **296**, L13–L16 (1995).
- Buzzoni, A. Evolutionary population synthesis in stellar systems. I. A global approach. *Astrophys. J. (Suppl. Ser.)* **71**, 817–870 (1989).
- Baraffe, I. & Chabrier, G. Mass-spectral class relationship for M dwarfs. *Astrophys. J.* **461**, L51–L54 (1996).
- Baraffe, I. et al. New evolutionary tracks for very low mass stars. *Astrophys. J.* **446**, L35–L38 (1995).
- Elson, R. A. W. et al. HST observation of the stellar population of the globular cluster ω Cent. *Astron. J.* **110**, 682–692 (1994).
- Mould, J. R. et al. Low mass stars in an outer field of NGC 6397. *Publ. Astron. Soc. Pacif.* **108**, 682–687 (1996).
- De Marchi, G. & Parsec, F. Low mass stars in globular cluster II. The mass function of M 15; Low mass stars in globular cluster III. The mass function of 47 Tucanae. *Astron. Astrophys.* **304**, 202–210; 211–219 (1995).
- Paresce, F., De Marchi, G. & Romaniello, M. Very low mass stars and white dwarfs in NGC 6397. *Astrophys. J.* **440**, 216–226 (1995).
- Cool, A. M., Pionto, G. & King, I. R. The main sequence and a white dwarf sequence in the globular cluster NGC 6397. *Astrophys. J.* **468**, 655–662 (1996).
- Hu, E. M., Huang, J.-S., Gilmore, G. & Cowie, L. L. An upper limit on the density of low-mass stars in the galactic halo. *Nature* **371**, 493–495 (1994).
- Bahcall, J. N., Flynn, C., Gould, A. & Kirhakos, S. M dwarfs, microlensing, and the mass budget of the Galaxy. *Astrophys. J.* **435**, L51–L54 (1994).
- Graff, D. S. & Freese, K. The mass function of low-mass halo stars: limits on baryonic halo dark matter. *Astrophys. J.* **467**, L65–L68 (1996).
- Flynn, C., Gould, A. & Bahcall, J. N. Hubble deep field constraints on baryonic dark matter. *Astrophys. J.* **466**, L55–L58 (1996).
- Frogel, J. A., Persson, S. E., Aaronson, M. & Matthews, K. Photometric studies of composite systems. I. CO and JHK observations of E and S0 galaxies. *Astrophys. J.* **220**, 75–97 (1978).
- Barnaby, D. & Thronson, H. A. The distribution of light in galaxies: the edge-on spiral NGC 5907. *Astron. J.* **103**, 41–53 (1992); Dark matter around galaxies: the edge-on spiral NGC 5907. *Astron. J.* **107**, 1717–1729 (1994).
- Alcaino, G., Liller, W. & Alvarado, F. BVRI CCD photometry of the globular cluster NGC 3201. *Astron. Astrophys.* **216**, 68–79 (1989).
- Alcaino, G. et al. BVRI CCD photometry of the globular cluster NGC 1904 (M79). *Astron. J.* **107**, 230–239 (1994).
- Sarajedini, A. & Milone, A. A. CCD photometry of NGC 5053: the most metal-poor galactic globular cluster. *Astron. J.* **109**, 269–279 (1994).
- Samus, N. et al. A CCD BVRI colour-magnitude study of the globular cluster NGC 5286. *Astron. Astrophys. Suppl. Ser.* **112**, 439–444 (1995).
- Ferraro, F. R. et al. Deep infrared array photometry of galactic globular clusters: the main sequence of NGC 6171. *Astron. Astrophys.* **298**, 461–464 (1995).
- Saumon, D., Bergeron, P., Lunine, J. I., Hubbard, W. B. & Burrows, A. Cool zero metallicity stellar atmospheres. *Astrophys. J.* **424**, 333–344 (1994).
- Reed, B. C., Hesser, J. E. & Shawl, S. J. Intrinsic integrated UBVR colors of galactic globular clusters. *Publ. Astron. Soc. Pacif.* **100**, 548–567 (1988).
- Bruzual, G. & Charlot, S. Spectral evolution of stellar populations using isochrone synthesis. *Astrophys. J.* **405**, 538–553 (1993).

Acknowledgements. We thank Y. Dotan, R. Young, A. Bhatnagar, T. Nguyen, C. Ibscher, D. Mabry, T. Tessensohn and D. Roux for help in preparing for these observing sessions, R. Cantera, D. Ciardi and G. Van Belle for logistical support, P. Sackett for discussions which focused our observations and F. Allard for communicating calculated colours for metal-deficient stars. This work was supported by the Aerospace Sponsored Research program. C.E.W. acknowledges support from the NSF.

Photochemical energy storage in a spatially organized zeolite-based photoredox system

Milan Sykora & James R. Kincaid

Chemistry Department, Marquette University, Milwaukee, Wisconsin 53233, USA

Zeolites are often employed as organizational media or supports for entrapped or adsorbed transition-metal catalysts and photocatalysts^{1–6}. In such applications, the individual catalytic species have been associated with the framework structure of the zeolite in a purely statistical (randomized) arrangement. A synthetic strategy developed recently⁷ has shown how a much higher level of organization can be obtained, so pointing the way to the generation of systems in which two or more active components can be arranged—both spatially and in terms of reactivity—within the zeolite host to enhance the efficiency of a desired catalytic reaction. Here we describe an application of this approach to photochemical storage of light energy. Such an application requires efficient photoinduced charge transfer between donor and acceptor molecules to form long-lived charge-separated states: the competing thermal back electron transfer reaction must be minimized. This is achieved in our system by arranging the active components (donor, acceptor and a ‘sensitizing’ intermediate molecule) such that they occupy adjacent cages within the zeolite framework, and results in unprecedented levels of net charge-separation efficiency.

Zeolites are aluminosilicates whose three-dimensional structure is made up of corner-sharing SiO_4 and AlO_4 tetrahedra, with exchangeable cations (M^{n+}) occupying extra framework positions to neutralize charge^{1,2}. One of the most commonly employed materials, Y-zeolite, has a three-dimensional structure comprised of tetrahedrally arranged interconnecting supercages of ~ 13 Å internal diameter which are connected to each other by 12-membered ring openings having 7–8 Å “windows”; a depiction of three such supercages is given in Fig. 1a. To clarify later discussion, we note that the four supercages surrounding a given central cage are not adjacent to one another.

Lunsford and co-workers were the first to show that the familiar photosensitizer $\text{Ru}(\text{bpy})_3^{2+}$, where bpy is 2,2'-bipyridine, could be synthesized (and thereby entrapped) within the supercages of Y-zeolite via a ‘ship-in-the-bottle’ approach^{8,9}. Dutta and Turbeville¹⁰ later showed that upon irradiation of $\text{Ru}(\text{bpy})_3^{2+}$ -loaded Y-zeolite, containing saturating amounts of methylviologen (MV^{2+}), a blue colour (indicative of $\text{MV}^{+ \cdot}$) was observed. More recently, Dutta and Borja¹¹, adapting a strategy advocated for other systems^{12,13}, have shown that intrazeolitic reducing equivalents ($\text{DQ}^{+ \cdot}$) can be transferred to an excluded, solution phase, acceptor (PVS), whose reduction potential is lower than that of DQ^{2+} , though the estimated quantum yields for PVS^- were disappointingly low (5×10^{-4}). (Here PVS is propylviologen sulphonate, and DQ^{2+} is $\text{N,N}'$ -tetramethylene-2,2'-bipyridinium.)

In view of the evident potential such materials hold for catalysis and photocatalysis, we have focused on the development of synthetic methods for organizational elaboration of zeolite-entrapped catalytic assemblies. These efforts are based on previously reported^{14–16} detailed procedures for the preparation of zeolite-entrapped bis-polypyridine complexes of divalent ruthenium (RuL_2^{2+}) and related bis-heteroleptic, tris-ligated polypyridine complexes ($\text{RuL}_2\text{L}'^{2+}$) and thorough documentation of their spectroscopic and photophysical properties.

Included among the zeolitic materials which have been prepared are those such as $\text{Z-Ru}(\text{bpy})_n(\text{bpz})_{3-n}^{2+}$ and $\text{Z-Ru}(\text{bpy})_n(\text{pypz})_{3-n}^{2+}$,

where bpz is 2,2'-bipyrazine and pypz is 2,2'-pyridylpyrazine. These complexes, possessing chelate ligands which bear peripheral nitrogens, can potentially serve as useful precursors in a synthetic route for construction of more elaborately organized intrazeolitic catalytic assemblies. Thus, as had been previously shown by Lever and co-workers¹⁷, in solution, the peripheral heteroatoms of such ligands are susceptible to coordination by various transition-metal-complex reagents; a reaction which we have recently reported⁷ is applicable to intrazeolitic complexes. These metallated, zeolite-entrapped, tris-ligated complexes can be further treated with excess chelating agent, rupturing the peripheral M-N_{bpz} bond, to produce a second transition-metal complex which is trapped in the supercage adjacent to that containing the original bis-heteroleptic, tris-ligated complex. Specifically, the zeolite sample to be discussed here (the preparation is described in ref. 7) contains dyad units, wherein the two complexes ($\text{Ru}(\text{bpy})_2(\text{bpz})^{2+}$ and $\text{Ru}(\text{mmb})_3^{2+}$) are entrapped in neighbouring supercages (as shown in Fig. 1a). (Here mmb is 5-monomethyl-2,2'-bipyridine.) Photophysical measurements, including luminescent lifetimes, are useful in documenting the fact that the majority of the entrapped complexes are in adjacent cages; that is, the newly formed complex does not migrate to other supercages, but remains trapped in the cage adjacent to the precursor complex (for example, $\text{Ru}(\text{bpy})_2(\text{bpz})^{2+}$).

We now consider why such adjacent-cage assemblies are useful in photoinduced charge separations. The unacceptably low photochemical quantum yields obtained for the zeolite-based ternary system ($\text{Ru}(\text{bpy})_3^{2+}/\text{DQ}^{2+}/\text{PVS}$) described above reflects the combined inefficiencies of several key steps in the overall cycle. Although the inefficiency of charge transfer at the zeolite–solution interface is undoubtedly an important factor, the great tendency for the initial charge-separated pair ($\text{Ru}(\text{bpy})_3^{2+ \cdot} \text{DQ}^{+ \cdot}$) to undergo the back-electron transfer (BET) reaction is apparently an important obstacle

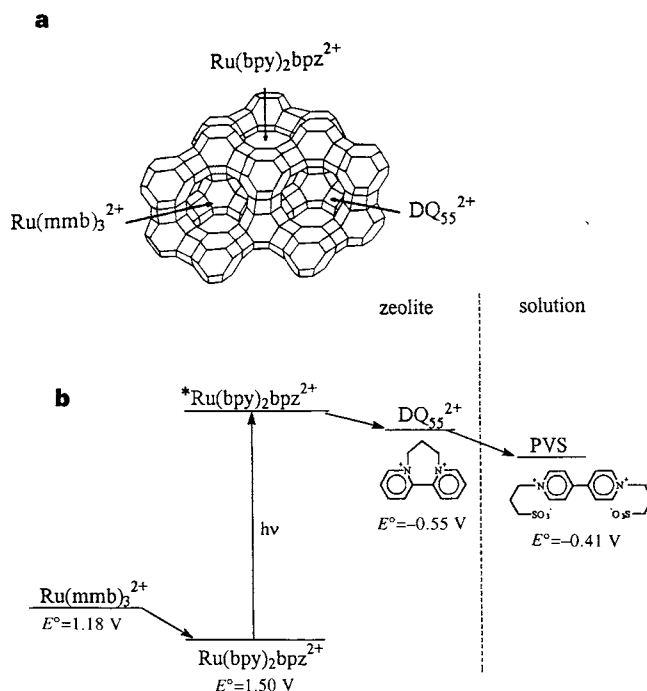


Figure 1 a, Schematic representation of three interconnected supercages of Y-zeolite and the arrangement of the donor ($\text{Ru}(\text{mmb})_3^{2+}$); sensitizer ($\text{Ru}(\text{bpy})_2(\text{bpz})^{2+}$); acceptor (DQ_{55}^{2+}) photocatalytic assembly. (DQ_{55}^{2+} is $\text{N,N}'$ -trimethylene-2,2'-bipyridinium, mmb is 5-monomethyl-2,2'-bipyridine, bpy is 2,2'-bipyridine, and bpz is 2,2'-bipyrazine.) b, Electron-transfer processes occurring inside the Y-zeolite particle, and at the zeolite-solution interface, on exposure to visible light. (PVS is propylviologen sulphonate.)

which plagues this system, as well as all other photoinduced charge-separation schemes^{3-5,11}. One strategy to overcome this limitation is to include a donor molecule in the photochemical cycle; that is, a donor which (in its oxidized form) retains oxidizing potential comparable to that of the oxidized sensitizer.

Referring to Fig. 1, the essential strategy being employed to attempt to increase the net charge-separation efficiencies of such systems is as follows. Similar to the original system employed by Dutta and co-workers^{10,11}, the photosensitizer is surrounded by acceptors (*N,N'*-trimethylene-2,2'-bipyridinium, DQ_{55}^{2+} , in the present case). Upon selective excitation (473 nm) of the entrapped $Ru(bpy)_2bpz^{2+}$ complex, an initial photoproduct ($Ru(bpy)_2(bpz)^{3+} \cdot DQ_{55}^{2+}$) is formed. Now, to the extent that reductive quenching of the oxidized sensitizer, $Ru(bpy)_2(bpz)^{3+}$, by the adjacent-cage $Ru(mmb)_3^{3+}$ is efficient and rapid, the resulting charge-separated pair ($Ru(mmb)_3^{3+}$ and DQ_{55}^{2+}) are now separated by an insulating (reduced) sensitizer. In the case where the $Ru(mmb)_3^{3+}$ component absorbs a photon, the dyad system is not expected to provide any advantage, inasmuch as the reduction of $Ru(mmb)_3^{3+}$ by $Ru(bpy)_2bpz^{2+}$ is thermodynamically quite unfavourable.

Shown in Fig. 2 are the diffuse reflectance spectra for the zeolite-entrapped dyads. As has been shown elsewhere¹⁴⁻¹⁶, the absorption bands for $Z-Ru(bpy)_2bpz^{2+}$ are centred at 412 and 473 nm, whereas that for $Z-Ru(mmb)_3^{3+}$ occurs near 446 nm. Although there is some overlap, absorption at 473 nm is mainly attributable to zeolite-

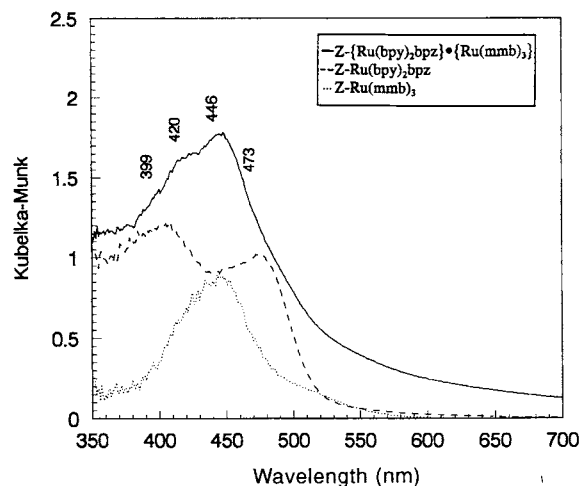
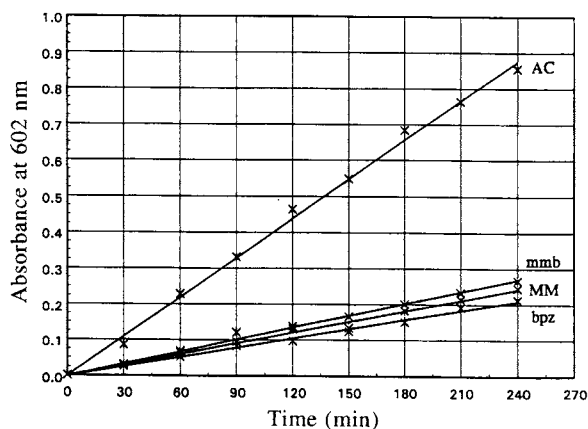


Figure 2 Diffuse reflectance spectra of the zeolite sample containing adjacent-cage dyads, $Z\{Ru(bpy)_2pbz^{2+} \cdot Ru(mmb)_3^{3+}\}$ (solid line); $Z-Ru(bpy)_2bpz^{2+}$ (dashed line); $Z-Ru(mmb)_3^{3+}$ (dotted line). Spectra were recorded in transmittance mode and subsequently numerically Kubelka-Munk corrected using facilities of Spectralcalc software.



entrapped $Ru(bpy)_2bpz^{2+}$. Shown in Fig. 3 are a series of absorption spectra acquired for the solution phase of a suspension containing the (DQ_{55}^{2+} -loaded) zeolite-entrapped dyads (later referred to as 'AC', that is, adjacent-cage) in a solution of PVS at various times during photolysis with 473-nm excitation. The spectra clearly document the continuous growth of solution-phase $PVS^{\cdot -}$. The magnitude of this growth, as a function of photolysis time, is plotted in Fig. 4.

Also plotted in Fig 4 is the growth profile for $PVS^{\cdot -}$ production for a reference system (MM) consisting of a suspension of a mixture of zeolitic materials in PVS solution. The reference system contains $Z-Ru(bpy)_2bpz^{2+}$ and $Z-Ru(mmb)_3^{3+}$, both of which are loaded with saturating levels of DQ_{55}^{2+} . It is emphasized here that both systems (the AC dyads and the MM reference sample) contain identical concentrations of both complexes and that the use of the MM reference ensures that for both systems the same number of $^*Ru(bpy)_2bpz^{2+}$ and $^*Ru(mmb)_3^{3+}$ excited states are created. Thus the increased efficiency of the adjacent-cage assemblies, relative to the reference system, is directly attributable to the unique spatial organization of the two complexes. The results indicate that the intrazeolitic dyads are approximately four times more efficient than the isolated sensitizers with respect to net charge separation. The results of photolysis studies on isolated systems, $Z-Ru(bpy)_2bpz^{2+}$ and $Z-Ru(mmb)_3^{3+}$ (total load of the complex in both cases being identical to the AC and MM systems) are also included in Fig. 4 and, as expected, are comparable to that for the MM reference system.

The efficiency of the charge-separation reaction in the original

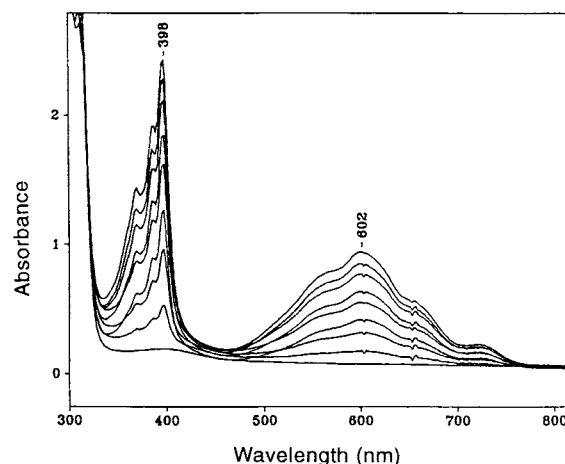


Figure 3 Changes in the absorption spectrum (every 30 min) of the solution phase on irradiation of the $Z\{Ru(bpy)_2pbz^{2+} \cdot Ru(mmb)_3^{3+}\}/1.0$ M PVS suspension with 100 mW of 473 nm (defocused) radiation from a Spectra-Physics 2025-05 Ar^+ laser.

Figure 4 Growth of the $PVS^{\cdot -}$ as a function of time for the adjacent-cage dyad assemblies, $Z\{Ru(bpy)_2pbz^{2+} \cdot Ru(mmb)_3^{3+}\}$ (AC), reference system (MM) composed of mechanically mixed $Z-Ru(mmb)_3^{3+}$ and $Z-Ru(bpy)_2bpz^{2+}$ and for the isolated systems, $Z-Ru(mmb)_3^{3+}$ (mmb) and $Z-Ru(bpy)_2bpz^{2+}$ (bpz). In all four systems the zeolite particles were loaded with identical amount of the Ru complex, that is $\sim 16 \mu\text{mol}$ of Ru complex per 1 g zeolite and then completely exchanged with *N,N'*-trimethylene-2,2'-bipyridinium (DQ_{55}^{2+}). The photolysis data for all four systems were obtained under identical conditions: 20 mg of zeolite sample was suspended in 0.1 M PVS solution, degassed and sealed anaerobically, then irradiated with 100 mW of 473-nm radiation. Care was taken to ensure that the stirring rates during the irradiation were identical for all systems. The solid lines represent the best linear least-squares fit to the experimental data. The calculated slopes are: $3.65 \times 10^{-3} \text{ min}^{-1}$ (AC/ DQ^{2+} /PVS); $1.01 \times 10^{-3} \text{ min}^{-1}$ (MM/ DQ^{2+} /PVS); $1.13 \times 10^{-3} \text{ min}^{-1}$ ($Z-Ru(mmb)_3^{3+}$ / DQ^{2+} /PVS), and $0.87 \times 10^{-3} \text{ min}^{-1}$ ($Z-Ru(bpy)_2pbz^{2+}$ / DQ^{2+} /PVS).

system^{10,11} (expected to be comparable to that of the present reference system) is not known with certainty¹¹. If it is assumed that a 20–30% yield of ‘permanent’ DQ⁺ results (that is, BET eliminates 70–80% of initially formed DQ⁺), then the ~four-fold increase in charge-separation efficiency observed here for the spatially organized system is all that could possibly be realized. That is, to the extent that the efficiency of the BET in the reference system is as low as 70–80%, the four-fold increase for the dyad system indicates that the strategy used here is indeed highly effective in reducing the effect of BET. We plan further work to document the BET reaction efficiency by time-resolved diffuse reflectance measurements and to attempt to improve the particle/solution-phase electron transfer process.

Here we have reported a specific application of the adjacent-cage assemblies to photoinduced charge-separation; we note that the synthetic methods which have been⁷ (and continue to be) developed should prove useful for a wide range of applications. Future efforts will be devoted to the construction of zeolite-entrapped assemblies consisting of redox-active complexes or photosensitizers coupled to coordinatively unsaturated, redox-active complexes which are capable of binding and activating small molecular substrates, such as the oxygen molecule and the oxides of carbon and nitrogen. □

Received 14 October 1996; accepted 25 March 1997.

1. Breck, D. W. *Zeolite Molecular Sieves* (Wiley, New York, 1974).
2. Weitkamp, J., Karge, H. G., Pfeifer, H. & Holderich, W. (eds) *Zeolites and Related Materials: State of the Art* (Elsevier, Amsterdam, 1994).
3. Kalyanasundaram, K. *Photochemistry in Microheterogeneous Systems* (Academic, New York, 1987).
4. O'Regan, B. & Grätzel, M. *Nature* **353**, 737–740 (1991).
5. Krueger, J. A., Meyer, J. E. & Mallouk, T. E. *J. Am. Chem. Soc.* **110**, 8232–8234 (1988).
6. De Vos, D. E. et al. *J. Incl. Phenom. Mol. Recogn. Chem.* **21**, 185–213 (1995).
7. Sykora, M., Maruszewski, K., Treffert-Ziemelis, S. M. & Kincaid, J. R. *J. Am. Chem. Soc.* (submitted).
8. DeWilde, W., Peters, G. & Lunsford, L. H. *J. Phys. Chem.* **84**, 2306–2310 (1980).
9. Quayle, W. H. & Lunsford, J. H. *Inorg. Chem.* **21**, 97–103 (1982).
10. Dutta, P. K. & Turbeville, W. J. *Phys. Chem.* **96**, 9410–9416 (1992).
11. Borja, M. & Dutta, P. K. *Nature* **362**, 43–45 (1993).
12. Willner, J., Yang, J., Loane, C., Otvos, J. W. & Calvin, M. J. *Phys. Chem.* **85**, 3277–3282 (1981).
13. Sliama-Schwork, A., Ottolenghi, M. & Avnir, D. *Nature* **355**, 240–242 (1992).
14. Maruszewski, K., Strommen, D. P., Handrich, K. & Kincaid, J. R. *Inorg. Chem.* **30**, 4579–4582 (1991).
15. Maruszewski, K., Strommen, D. P., Handrich, K. & Kincaid, J. R. *J. Am. Chem. Soc.* **115**, 8345–8350 (1993).
16. Maruszewski, K. & Kincaid, J. R. *Inorg. Chem.* **34**, 2002–2006 (1995).
17. Toma, H. E., Auburn, P. R., Dodsword, E. S., Golovin, M. N. & Lever, A. B. P. *Inorg. Chem.* **26**, 4257–4263 (1987).

Acknowledgements. This work was supported by the Division of Chemical Sciences, US Department of Energy.

Correspondence and requests for materials should be addressed to J.R.K. (e-mail: Kincaid@vms.csd.mu.edu).

Total synthesis of the potential anticancer vaccine KH-1 adeno carcinoma antigen

Prashant P. Deshpande* & Samuel J. Danishefsky*†

* Laboratory for Bioorganic Chemistry, Sloan-Kettering Institute For Cancer Research, 1275 York Avenue, New York, New York 10021, USA

† Department of Chemistry, Columbia University, Havemeyer Hall, New York, New York 10027, USA

Human tumours are often marked by the expression of unusual carbohydrate structural motifs^{1–3}. These carbohydrate domains are manifested as cell-surface bound glycolipids or glycoproteins⁴. This raises the possibility of using cell-free equivalents of these domain compounds, obtained by total synthesis with a view towards triggering some level of immune response. In fact, the serum of mice immunized with fully synthetic compounds^{5–7} that mimic cell-surface tumour antigens has already been shown to recognize pertinent human cancer cell lines⁸. Further advances in

this field depend critically on the availability of these tumour-associated carbohydrate antigens which cannot be readily isolated from natural sources in sufficient quantities. Here we present the successful total synthesis of an adenocarcinoma antigen, KH-1, and of a bioconjugatable analogue which can bind to a carrier protein. These results illustrate the capabilities of oligosaccharide synthesis for reconstructing the challenging structural motifs characteristic of carbohydrate antigens, and thereby open up new possibilities for the development of anticancer vaccines.

The chances of success from the point of view of vaccinology will be higher with the greater specificity of the carbohydrate domain of the antigen. From this perspective, one such tumour antigen which attracted our notice is the glycolipid KH-1, immunocharacterized by Hakomori and associates⁹. (We note that a total synthesis of KH-1 antigen, by a classical route, was reported¹⁰ after submission of this Letter.) This antigen is apparently a highly telling marker for malignancy and pre-malignancies involving colonic adenocarcinoma. Arguments have been advanced^{10–12} asserting the structure of KH-1 to be **1** (shown in Fig. 1). The nonasaccharide character of **1** is, from a structural standpoint, unique. In considering KH-1 as a candidate for total synthesis we also noted the crystallographically derived presentation of monoclonal antibody BR96, bound to a Le^y tetrasaccharide glycoside¹³. The structure of the Br96: Le^y complex suggested that this antibody might have the capacity to recognize higher-order fucosylated arrays. With a view to addressing this question, as well as investigating the possibility of incorporating KH-1 in a vaccine, we have directed considerable attention towards the total synthesis. Our inquiry was not limited to KH-1 (**1**), but included congeners^{14,15} (structure **2**; Fig. 1) which are suitable for conjugation to appropriate bioactive carrier systems. We report here the attainment of these goals.

In conducting this project, we also hoped to address the important issue of ‘strategy’ in oligosaccharide synthesis. Of course in this field, as opposed to ‘conventional’ natural product synthesis, the basic building blocks are rather restricted and tend to bear obvious homology with readily recognized components of the target system. Nonetheless, as we show, there are considerable opportunities in the design of convergent strategies, which might lead to synthetic economies.

From this perspective we came to favour a plan which would build a hexasaccharide (structure **13** in Fig. 3), so differentiated in terms of its protecting patterns (see arrows) as to allow for the unveiling of the three free hydroxyls to serve as α fucosylation acceptor sites. In this way, perhaps, the three immunologically defining α -fucose units could be introduced in one concurrent synthetic operation. Considerable thought was also directed to assembling the hexasaccharide with sound management of an otherwise potentially forbidding network of hydroxyl group functionalities. We found advantages in drawing from principles which emerged from the logic of glycal assembly¹⁶.

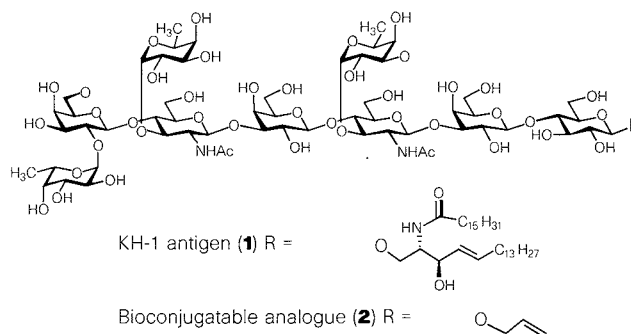


Figure 1 The structure of KH-1 antigen and a congener that is suitable for conjugation to appropriate bioactive carrier systems.

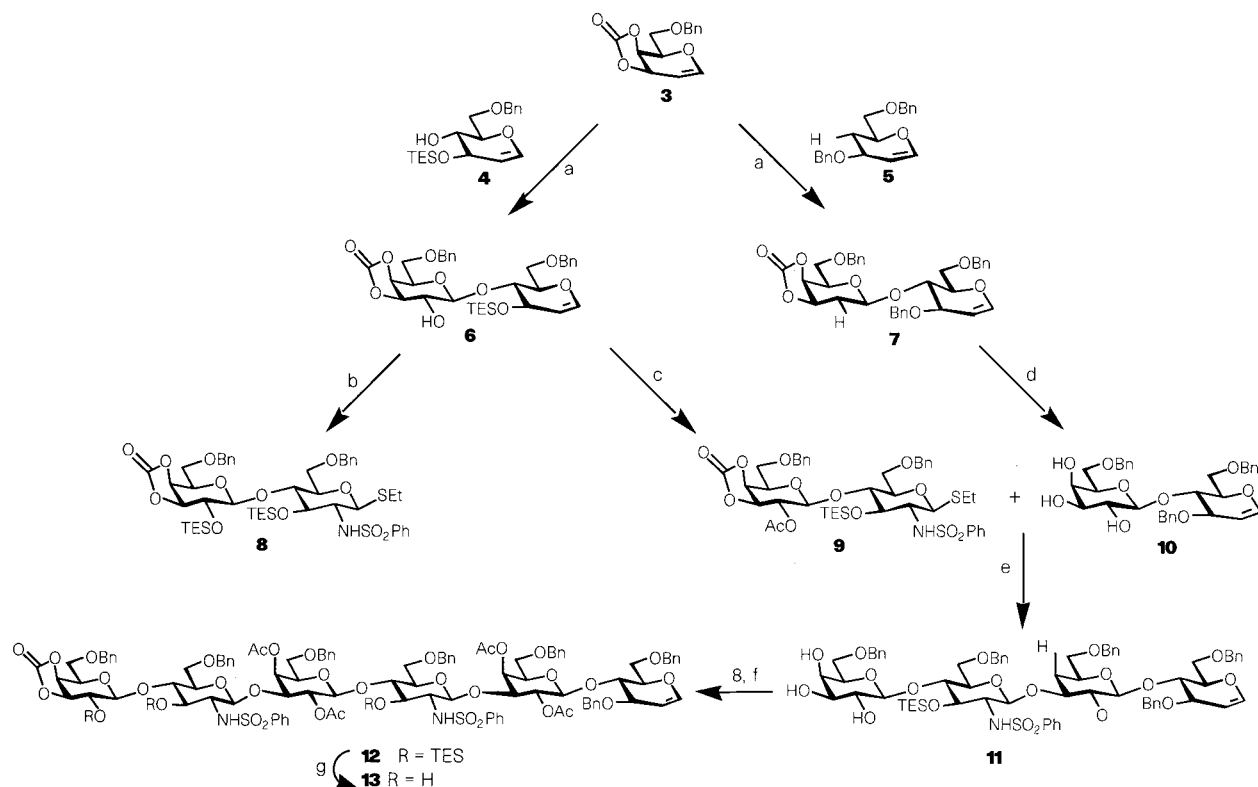


Figure 2 Reagents: a, (i) 3,3-dimethyldioxirane, CH_2Cl_2 , (ii) **4** or **5**, ZnCl_2 , THF 65% for **6** and 55% for **7**; b, (i) TESOTf, Et_3N , CH_2Cl_2 92%, (ii) $\text{I}(\text{coll})_2\text{ClO}_4$, PhSO_2NH_2 , 4-Å molecular sieves (MS), CH_2Cl_2 , >90%, (iii) LHMDs, EtSH, DMF >90%; c, (i) Ac_2O , Et_3N , DMAP, CH_2Cl_2 , 95%, (ii) $\text{I}(\text{coll})_2\text{ClO}_4$, PhSO_2NH_2 , 4-Å MS, CH_2Cl_2 , >90%, (iii) LHMDs, EtSH, DMF, (iv) Ac_2O , Et_3N , DMAP, CH_2Cl_2 , 85%; d, K_2CO_3 , MeOH 80%; e, (i) MeOTf, di-*t*-butylpyridine, $\text{Et}_2\text{O}:\text{CH}_2\text{Cl}_2$ (2:1), 4-Å MS (55%), (ii) K_2CO_3 , MeOH

(85%); f, (i) MeOTf, di-*t*-butylpyridine, $\text{Et}_2\text{O}:\text{CH}_2\text{Cl}_2$ (2:1), 4-Å MS (60%), (ii) Ac_2O , Py, DMAP, CH_2Cl_2 (95%); g, TBAF:AcOH (93%). (THF, tetrahydrofuran; TESOTf, triethylsilyl trifluoromethanesulphonate; $\text{I}(\text{coll})_2\text{ClO}_4$, Bis-*sym*-collidine iodonium perchlorate; LHMDs, lithium bis(trimethylsilyl)amide; DMF, *N,N*-dimethylformamide; DMAP, *N,N*-dimethylaminopyridine; MeOTf, methyl trifluoromethanesulphonate; TBAF, tetrabutylammonium fluoride; OBn, O-benzyl.)

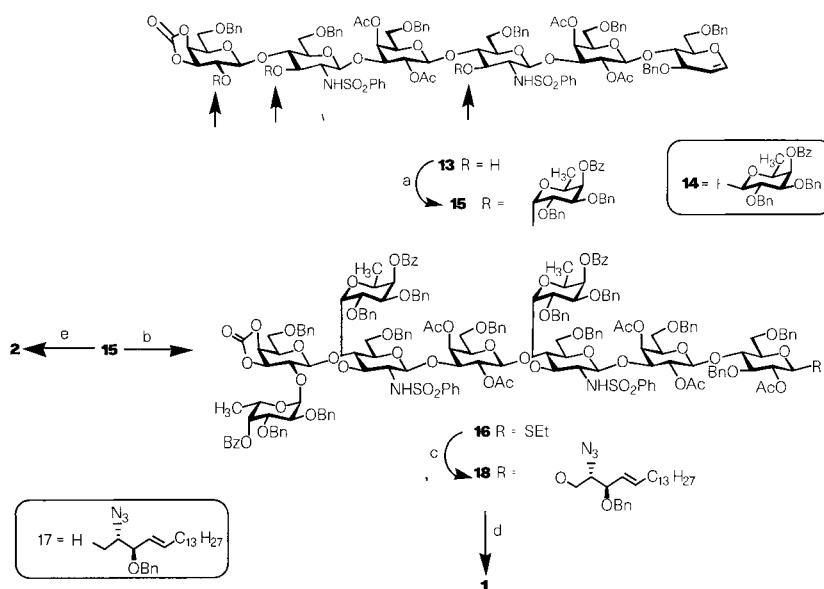


Figure 3 a, **14**, $\text{Sn}(\text{OTf})_2$, toluene:THF(10:1), 4-Å molecular sieve (MS) (60%); b, (i) 3,3-dimethyldioxirane, CH_2Cl_2 , (ii) EtSH, CH_2Cl_2 , H^+ (cat.), (iii) Ac_2O , Py, CH_2Cl_2 60% (three steps); c, **17**, MeOTf, $\text{Et}_2\text{O}:\text{CH}_2\text{Cl}_2$ (2:1), 4-Å MS (55%); d, (i) Lindlar's catalyst, H_2 , palmitic anhydride, EtOAc, 85%, (ii) Na^0NH_3 , THF, (MeOH quench),

(iii) Ac_2O , Et_3N , DMAP, CH_2Cl_2 , (iv) MeONa, MeOH, 70% (three steps); e, (i) Na^0 , NH_3 , THF; (MeOH quench), (ii) Ac_2O , Et_3N , DMAP, CH_2Cl_2 , (iii) 3,3-dimethyldioxirane, CH_2Cl_2 , (iv) allyl alcohol, (v) MeONa, MeOH, 60%. (H^+ (cat), catalytic acid; Na^0 , sodium; OBz, O-benzoyl.)

Thus differentiated glycals **4** (ref. 17) and **5** were derived from D-glucal by exploiting the reliable reactivity preference of the C₆, C₃ and C₄ hydroxyls (C₆ > C₃ > C₄)^{18,19}; compounds **3–13** are shown in Fig. 2). Moreover, the fashioning of a clean α-epoxide from galactal derivative **3** is well known¹⁶, as is the excellent β-galactosyl donating capacity of such an epoxide²⁰. Coupling of this epoxide to acceptors **4** and **5** under mediation by anhydrous zinc chloride gave **6** and **7** respectively. The C₂' hydroxyl of the lactal derivative **6** was silylated and the resultant glycal linkage was subjected to overall suphonamido (2α) ethanethiolation^{3,21} (1β) to provide **8**. We note that in this product two of the three sites destined for eventual fucosylation have been distinguished. In a parallel experiment, the acetate derived from compound **6** was converted to intermediate **9** which carried the third eventual fucosylation centre at the site of its triethylsilyl (TES) protecting group. Cleavage of the carbonate linkage of **7** generated triol **10**. Here, we could take advantage of another well appreciated preference^{22,23} which is that the glycosyl accepting site in such a triol tends to be at the C₃' hydroxyl group. Coupling of **10** and **9** afforded, after cleavage of the cyclic carbonate and acetate linkages, a pentaol (structure **11**).

At this stage, we chanced the proposition that the 1, 2, 3 triol in the terminal ring, rather than the 1, 3 diol in the galactose ring adjacent to glycal would serve as the acceptor locus with donor **8**. This in fact proved to be the case. The successful glycosylation was followed by acetylation of the four remaining hydroxyl groups. This sequence led to **12** and thence to **13** containing the three free hydroxyl sites targeted for fucosylation.

It proved possible to introduce the three α-L-fucose residues in one step via donor **14** (ref. 24), thereby affording a 60% yield of the nonasaccharide **15**; compounds **14–18** are shown in Fig. 3. From this point, the protocols required to reach **1** and **2** were much influenced by our earlier work in the globo H series^{5–7}. The application of this chemistry to reach the pre-conjugation allyl glycoside **2** is provided in Fig. 3. To reach the naturally occurring glycolipid antigen **1**, we introduced a useful variation wherein the preceramide acceptor **17** was coupled to an anomeric thioethyl donor derived from the glycal epoxide (see **15** → **16**)^{25,26} to generate compound **18**.

The structures of the final products **1** and **2** were fully substantiated by mass spectroscopy, by self-consistent NMR analysis and, in the case of **1**, by correspondence with the available fragmentary published data⁵. A direct material comparison was not possible because of the nonfeasibility of obtaining a sufficient quantity of KH-1 antigen from natural sources. Owing to the synthesis reported here, access to this system is no longer an impassable problem. □

Received 3 February; accepted 2 April 1997.

- Hakomori, S. Aberrant glycosylation in cancer cell membranes as focused on glycolipids: overview and perspectives. *Cancer Res.* **45**, 2405–2414 (1985).
- Feizi, T. Carbohydrate antigens in human cancer. *Cancer Surv.* **4**, 245–269 (1985).
- Lloyd, K. O. Humoral immune responses to tumor-associated carbohydrate antigens. *Cancer Biol.* **2**, 421–431 (1991).
- Hakomori, S. New directions in cancer therapy based on aberrant expression of glycosphingolipids: anti-adhesion and ortho-signaling therapy. *Cancer Cells* **3**, 461–470 (1991).
- Park, T.-K. *et al.* Total synthesis and proof of structure of a human breast tumor (globo-H) antigen. *J. Am. Chem. Soc.* **118**, 11488–11500 (1996).
- Bilodeau, M. T. *et al.* Total synthesis of a human breast tumor associated antigen. *J. Am. Chem. Soc.* **117**, 7840–7841 (1995).
- Kim, I. J. *et al.* Defining the molecular recognition of globo H (human breast cancer) antigen through probe structures prepared by total synthesis. *J. Org. Chem.* **60**, 7716–7717 (1995).
- Ragupathi, G. *et al.* Immunization of mice with a fully synthetic globo H antigen results in antibodies against human cancer cells: a combined chemical-immunological approach to the fashioning of an anticancer vaccine. *Angewandte Chemie* **36**, 125–128 (1997).
- Nudelman, E., Levery, S. B., Kaizu, T. & Hakomori, S.-I. Novel fucolipids of human adenocarcinoma: characterization of the major Le^x antigen of human adenocarcinoma as trifucosylmannosyl Le^x glycolipid (III²FucV³FucVI³FucnI₆). *J. Biol. Chem.* **261**, 11247–11253 (1986).
- Hummel, G. & Schmidt, R. R. A versatile synthesis of the Lactoneo-series antigens-synthesis of sialyl dimer Lewis X and dimer Lewis Y. *Tetrahedron Lett.* **38**, 1173–1176 (1997).
- Kaizu, T., Levery, S. B., Nudelman, E., Stenkamp, R. E. & Hakomori, S.-I. Novel fucolipids of human adenocarcinoma: monoclonal antibody specific for trifucosyl Le^x (III²FucV³FucVI³FucnI₆) and a possible three-dimensional epitope structure. *J. Biol. Chem.* **261**, 11254–11258 (1986).
- Kim, S. Y. Expression of Le^x and extended Le^x blood group-related antigens in human malignant, premalignant, and nonmalignant colonic tissues. *Cancer Res.* **46**, 5985–5992 (1986).
- Jeffrey, P. D. The X-ray structure of an anti-tumour antibody in complex with antigen. *Nature Struct. Biol.* **2**, 466–471 (1995).

- Helland, A.-C., Nilsson, M. & Norberg, T. Synthesis of a trifucosyl Le^x heptasaccharide corresponding to a tumor-associated glycolipid. *J. Carbohydr. Chem.* **11**, 77–88 (1992).
- Wildmüller, R. & Schmidt, R. R. Efficient synthesis of Lactoneo series antigens H, Lewis X (Le^x), and Lewis Y (Le^y). *Tetrahedron Lett.* **35**, 7927–7930 (1994).
- Danilshesky, S. J. & Bilodeau, M. T. Glycals in organic synthesis: the evolution of comprehensive strategies for the assembly of oligosaccharides and glycoconjugates of biological consequences. *Angew. Chem. Int. Edn Engl.* **35**, 1380–1419 (1996).
- Blackburne, I. D., Fredericks, P. F. & Guthrie, R. D. Studies on unsaturated sugars with particular reference to the synthesis of 6-deoxy-6-fluoro derivatives. *Aust. J. Chem.* **29**, 381–391 (1976).
- Danilshesky, S. J. *et al.* Application of glycals to the synthesis of oligosaccharides: convergent total syntheses of the Lewis X trisaccharide sialyl Lewis X antigen determinant and higher congeners. *J. Am. Chem. Soc.* **117**, 1940–1953 (1995).
- Danilshesky, S. J., Behar, V., Randolph, J. T. & Lloyd, K. O. Application of the glycal assembly method to the concise synthesis of neoglycoconjugates of Le^x and Le^y blood group determinants of H-type I and H-type II oligosaccharides. *J. Am. Chem. Soc.* **117**, 5701–5711 (1995).
- Gervay, J., Peterson, J. M., Oliyama, T. & Danilshesky, S. J. An unexpected sialylation: total syntheses of G_{M1} and a positional isomer. *J. Org. Chem.* **58**, 5465–5468 (1993).
- Griffith, D. A. & Danilshesky, S. J. Sulfonamidoglycosylation of glycals. A route to oligosaccharides with 2-aminohexose subunits. *J. Am. Chem. Soc.* **112**, 5811–5819 (1990).
- Kameyama, A., Ishida, H., Kiso, M. & Hasegawa, A. Synthetic studies on sialoglycoconjugates 22: total synthesis of tumor-associated ganglioside, sialyl Lewis X. *J. Carbohydr. Chem.* **10**, 549–560 (1991).
- Kondo, T., Tomoo, T., Abe, H., Isobe, M. & Goto, T. Total synthesis of GD₃, a ganglioside. *Chem. Lett.* **5**, 337–338 (1996).
- Danilshesky, S. J. *et al.* Remarkable regioselectivity in the chemical glycosylation of glycal acceptors: a concise solution to the synthesis of sialyl Lewis X glycal. *J. Am. Chem. Soc.* **114**, 8329–8331 (1992).
- Fügedi, P., Garegg, P. J., Lönn, H. & Norberg, T. Thioglycosides as glycosylating agents in oligosaccharide synthesis. *Glycoconjugate J.* **4**, 97–108 (1987).
- Lönn, H. Glycosylation using a thioglycoside and trifluoromethanesulfonate. A new and efficient method for cis and trans glycoside formation. *J. Carbohydr. Chem.* **6**, 301–306 (1987).

Acknowledgements. We thank G. Sukenick for NMR and mass spectral analyses and D. Live for NMR analyses. We also thank T.-K. Park for suggestions about the strategy of the synthesis. This work was supported by the National Institutes of Health.

Correspondence should be addressed to S.J.D.

A major biopolymeric component to dissolved organic carbon in surface sea water

Lihini I. Aluwihare*, Daniel J. Repeta* & Robert F. Chen†

* Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole Massachusetts 02543, USA

† Environmental and Coastal Oceans Sciences Program, University of Massachusetts, Boston, 100 Morrissey Boulevard, Boston, Massachusetts 02125-3393, USA

Organic carbon dissolved in sea water is an important component of the global carbon cycle¹. Concentrations of dissolved organic carbon (DOC) in the ocean's surface mixed layer are at least twice those in the deep sea^{2,3}, because of the production of soluble carbon compounds by marine algae in the euphotic zone^{4,5}. But very little is known about the chemical composition of DOC, and the connection between photosynthetic production and DOC accumulation is not well understood^{6,7}. Here we report the chemical characterization of macromolecular DOC at several sites in the Atlantic and Pacific oceans. Neutral sugars, acetate and lipids show similar distributions, suggesting that these constituents are linked together in a common macromolecular structure. Chemical linkage patterns between the oligosaccharide portions of dissolved organic matter subjected to ultrafiltration are highly specific, with little variation between ocean basins. We show that laboratory culture experiments on the decomposition of algal exudate produce macromolecular organic matter with similar compositions and linkage characteristics. We propose that a significant fraction of DOC in sea surface water consists of structurally related and biosynthetically derived acyl oligosaccharides that persist after more labile organic matter has been degraded.

Between 25% and 35% of marine DOC is in a high-molecular-weight fraction that can be recovered by ultrafiltration or dialysis^{8–10}.

Table 1 Sample location, DOC concentration, and the relative abundance of major biochemicals in UDOM

Sample* (date)	Location	DOC (μM)	Carbon (relative %)		
			Carbohydrate†	Acetate	Lipid‡
Atlantic Ocean					
Georges Bank(Mar. 93)					
GB-1	40° 30' N, 70° 45' W	80	81	11	4
GB-2	40° 58' N, 68° 54' W	85	76	9	14
Mid-Atlantic Bight (Apr. 94)					
MAB-1	40° 32' N, 72° 09' W	95	73	10	15
MAB-5	38° 56' N, 75° 05' W	116	84	6	10
MAB-6	37° 43' N, 75° 24' W	99	75	9	16
MAB-7	37° 11' N, 74° 16' W	97	77	10	14
Woods Hole (Jul. 94)					
WH-1	41° 32' N, 70° 31' W	102	86	10	4
WH-2‡	41° 32' N, 70° 31' W	102	77	15	7
Oosterschelde (Apr.-Jun. 94)§					
Os-1	51° 36' N, 04° 07' E	260	84	11	4
Pacific Ocean					
Scripps Pier (Jun. 93)					
SP-1	32° 52' N, 117° 18' W	ND¶	81	12	6
Peru Upwelling (Nov. 92)					
PU-1	11° 04' S, 78° 04' W	ND	81	13	6
Hawaii (Jan. 95)	19° 40' N, 156° W	ND	85	7	8
Average composition (all samples)			80 ± 4	10 ± 2	9 ± 4

* Samples from Georges Bank, the Mid-Atlantic Bight, Woods Hole, Scripps Pier, Hawaii and Oosterschelde were collected with either an Amicon DC-10 or Filtron ultrafiltration system fitted with a 1 K membrane. Samples were collected in all-Teflon system with cartridge filtration at 0.2 μm. Water was processed immediately with collection. Concentrates were de-salted by diafiltration with deionized (Milli-Q) water. Carbon was monitored in feed, retentate and permeate, and mass balances were ± 15%. All samples were collected between 1–15 m. Water from the Peru Upwelling area was collected at 100 m, frozen, and returned to the laboratory where the > 1 K fraction was isolated by dialysis against deionized water using a cellulose membrane. † Carbohydrate, acetate and lipids were determined by ¹H NMR after normalization for differences in C/H (carbohydrates 1; acetate 0.67; lipids 2.5), and assuming 30% deoxysugar content. ‡ Values are expressed relative to the summed total. NMR spectra were acquired on a Bruker AC-300 (MHz) spectrometer in D₂O, using water suppression. § > 10 K sample. ¶ Composite sample collected between 25 April and 8 June. || Ref. 19. ¶ Not reported/determined.

Using these techniques, organic matter was sampled at sites in the Atlantic and eastern tropical Pacific oceans (Table 1). Proton NMR of ultrafiltered dissolved organic matter (UDOM) yields very similar spectra for all samples, with well resolved resonances from carbohydrates (5.5–5 p.p.m. (anomeric), 4.3–3.4 p.p.m. (CH), and 1.3 p.p.m. (CH₃); 55% total carbon), acetate (2.0 p.p.m.; 7% total carbon), and low-molecular-weight lipids (1.3 (CH₂) and 0.9 (CH₃) p.p.m.; 6% total carbon) (Fig. 1a). Carbohydrate, acetate and lipids determined by ¹H NMR are in good agreement with ¹³C NMR data for Woods Hole sea water (L.I.A. and D.J.R., unpublished results) and for the North Pacific Ocean⁸. The abundance of major biochemicals in UDOM is relatively constant in all samples, and has an average carbohydrate/acetate/lipid carbon ratio of 8/1/1 (Table 1). Compositional similarities of UDOM are also evident in the monomer components of the carbohydrate. Monosaccharide analysis of hydrolysed UDOM in six samples spaced across the US Mid-Atlantic Bight, between Cape Cod and Cape Hatteras (WH-1, MAB 1–7), Hawaii, and in the eastern North Atlantic (Oosterschelde) show a fixed ratio of major neutral sugars (Fig. 2). Galactose (0.20 ± 0.03 mol%) is the most abundant monosaccharide in the samples. Xylose (0.12 ± 0.02), rhamnose (0.16 ± 0.02), fucose (0.15 ± 0.1), glucose (0.16 ± 0.03) and mannose (0.13 ± 0.03) are only slightly less abundant. Arabinose (0.07 ± 0.02), the least abundant sugar, is 35% of the concentration of galactose. This complex distribution of neutral sugars is similar to UDOM neutral sugar distributions in the Gulf of Mexico, Sargasso Sea, and North Pacific Ocean (Fig. 2)^{11,12}. Data presented in Table 1 and Fig. 2 span a range of oceanographic settings, geographical areas and collection dates. DOC concentrations vary threefold, from 80 μM to 260 μM, at the collection sites.

Polysaccharide linkage analysis¹³ of UDOM recovered from the eastern and western North Atlantic and Pacific oceans was used to investigate the degree of polymerization and crosslinking of the carbohydrate fraction. Quantitatively, 40% of all sugars have terminal linkages, 40% are linked without branching, and 20% are

crosslinked with one branch point. The ratio of terminal/non-branched/branched linkages yields an average polymer consisting of five sugars (DP₅), having a minimum (nominal) relative molecular mass of 1,000 (1K), has one branch point, and is acetylated at two sites (carbohydrate/acetate molar ratio of 5:2). Algal polysaccharides are synthesized with a wide range of linkages, including (1,2), (1,3), (1,4) and (1,6) for non-branched sugars, and (1,2,3), (1,2,4), (1,3,4), (1,3,6), (3,4,6) for branched sugars. However, within specific polysaccharides, a much smaller number of linkages are observed¹⁴. UDOM is likewise characterized by a limited number of specific linkages and is highly branched. A fraction of all sugars have terminal linkages (Table 2). Only a few sugars are branched, and with a single exception (3, 4, 6 galactose in the Oosterschelde sample), only 1, 3, 4 branching is observed. For non-branched sugars, major linkages include 1, 3 arabinose (100%), 1,3 rhamnose (50–80%), 1,3 fucose (90–100%), 1,4 xylose (30–80%), 1,2 mannose (40–100%), 1,3 glucose (50–100%) and 1,3 galactose (90–100%).

The relatively fixed ratio of major biochemicals and simple oligosaccharide linkage patterns observed in UDOM suggests that a large fraction of macromolecular DOC in surface sea water is not the complex, heterogeneous polymers expected from geopolymerization of simple biomolecules^{15,16}. Rather, our results show that most macromolecular organic matter is a mixture of structurally related acyl-oligosaccharides. Carbohydrate, acetate and lipid portions of UDOM are linked together in a common macromolecular structure, giving rise to the relatively fixed ratio of major biochemicals measured by our analyses. The carbohydrate portion of this macromolecule is also characterized by a distinctive and very heterogeneous distribution of seven neutral sugars. Processes responsible for the formation of these macromolecules must be operative on a global scale. Observed variations in monosaccharide distributions, and in the ratio of carbohydrate, acetate and lipid, may result from differences in UDOM sources and removal processes between our different sampling sites. Indeed, UDOM

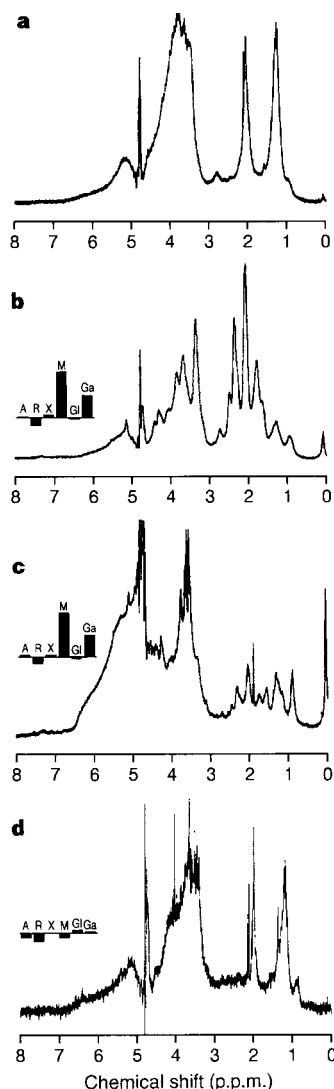


Figure 1 **a**, Proton NMR of UDOM from sample Woods Hole sea water showing major resonances from carbohydrate (5.5–5.0 (anomeric), 4.3–3.4 (CH) and 1.3 p.p.m. (CH₃ from deoxysugars), acetate (2.0 p.p.m.), and low-molecular-weight lipids (1.3 and 0.9 p.p.m.). These resonances from major biochemicals are superimposed on a featureless background of overlapping resonances that accounts for the remaining 30% carbon in our samples. Aromatic constituents (6.8–8.0 p.p.m.) contribute <1% of the total carbon in our samples, with slightly higher values observed in coastal seawater samples that are heavily influenced by freshwater inputs from nearby estuaries. Acetate and lipids are not extractable from UDOM (redissolved in water, pH 1) with organic solvents. However, after hydrolysis with 4 M HCl (90 °C; 30 h), acetate and lipids are recoverable in the organic extract. The identification of acetate was confirmed by ion chromatography, which yielded 94% of the expected amount measured by NMR. **b**, NMR spectrum, collected immediately after algae were removed from the culture, showing major resonances from carbohydrates, proteins and lipids. Monosaccharide distributions (inset); where A is arabinose, R is rhamnose, X is xylose, M is mannose, Gl is glucose and Ga is galactose) are given as the difference between culture and average seawater data (Fig. 2) normalized to fucose. Positive deviations correspond to a higher monosaccharide relative abundance in the culture sample. **c**, NMR spectrum showing the selective degradation of dissolved proteins (2.8–1.5 p.p.m.) after 19 d incubation. **d**, After 37 d, the distribution of major biochemicals and monosaccharides is similar to UDOM collected from sea water (**a**). An incubated control sample of permeate (<1K DOC) without added algae shows no significant UDOM fraction, and a control sample poisoned (HgCl₂) after the removal of algae likewise shows no change with incubation from NMR spectrum **b**. Samples were dissolved in D₂O, and chemical shifts are expressed relative to water at 4.8 p.p.m.

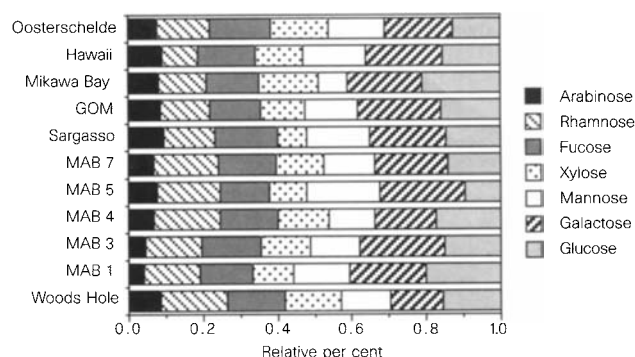


Figure 2 Normalized distribution of monosaccharides in UDOM samples collected from Woods Hole, the Mid-Atlantic Bight (MAB), Hawaii and Oosterschelde. Data from Gulf of Mexico (GOM), Sargasso Sea (Sargasso), and Mikawa Bay surface water are given for comparison^{11,12}. Gluco- and galactouronic acid and amino sugar (not shown) distributions are more variable than neutral sugars, but contribute <12% of total monosaccharides in our samples. Data were generated for unpurified samples using a variety of analytical methods for hydrolysis and monomer quantitation. The comparison therefore includes differences in the data due to the presence of other UDOM constituents and the inherent biases of the respect analytical techniques. Sample locations for MAB-3 and MAB-4 are 37° 40' N, 73° 25' W and 39° 16.6' N, 75° 21.2' W respectively.

carbohydrates include the same monosaccharides identified in transparent exopolysaccharides from marine snow, suggesting one possible removal mechanism¹⁷. Acyl-oligosaccharides in UDOM are probably formed by direct biosynthesis, and persist after the removal of more labile constituents.

To simulate the production and accumulation of the oligosaccharide in sea water, we conducted a simple laboratory experiment to monitor changes in UDOM during decomposition of algal exudate. Water (180 l) collected from 750 m depth (DOC, 45 µM) at a site offshore of Cape Hatteras, was filtered to remove particles, and ultrafiltered to remove UDOM. Nutrients (N, P, Si), trace metals, vitamins, and a seed culture of the marine diatom *Thalassiosira weissflogii* were added to the permeate (<1K) which was then incubated under artificial light at room temperature. After the algae reached stationary phase growth, and DOC values began to decline, the culture was filtered to remove algae and incubated again in the dark. After 19 days, 21 of Vineyard Sound sea water (pre-screened to 53 µM, Woods Hole, MA) was added to inoculate the sample with a natural consortium of bacteria and protozoa.

The sequence of NMR spectra in Fig. 1b–d traces the production of dissolved organic matter and the accumulation of acyl-oligosaccharides. At the onset of stationary phase growth, DOC concentrations had reached 430 µM. The NMR spectrum of UDOM (Fig. 1b) shows major contributions from polysaccharides (4.3–3.2 p.p.m.), proteins (2.8–1.5 p.p.m.) and minor amounts of lipid (1.5–0.8 p.p.m.). After 19 days of degradation, the amounts of major biochemicals changed considerably (Fig. 1c); sugars and lipids are relatively more abundant, but a significant protein fraction remains. Finally, 18 days after the addition of Vineyard Sound sea water (37 days total incubation), the NMR spectrum of UDOM (Fig. 1d) is similar to UDOM in sea water (Fig. 1a). The monosaccharide composition of our first incubation sample is dominated by mannose (42%) and galactose (26%), but all other major sugars measured in UDOM are present (Fig. 1b–d, inset). After 19 days, the relative abundances of mannose and galactose were unchanged (41% and 27%, respectively). After 37 days, mannose and galactose decreased to respectively 8% and 21% of the total sugar, and the ratio of major neutral sugars approached UDOM values. Oligosaccharide linkage patterns at the end of our experiment were highly specific and nearly identical to UDOM isolated from sea water (Table 1).

Table 2 Oligosaccharide linkages for culture and surface seawater UDOM

	Culture			Oosterschelde			Hawaii			Woods Hole		
	T*	B	NB	T	B	NB	T	B	NB	T	B	NB
Arabinose (f)†	100	0	0	100	0	0	100	0	0	100	0	0
Arabinose (p)‡	25	33 (1, 3, 4)	42 (1, 3)	31	0	68 (1, 3)	18	48 (1, 3, 4)	34 (1, 3)	55	0	45 (1, 3)
Rhamnose	22	4 (1, 3, 4)	43 (1, 3) 27 (1, 4)	54	9 (1, 3, 4)	22 (1, 3) 17 (1, 4)	60	27 (1, 3, 4)	10 (1, 3) 3 (1, 4)	78	0	11 (1, 3) 11 (1, 4)
Fucose	36	24 (1, 3, 4)	12 (1, 3) 26 (1, 4)	34	29 (1, 3, 4)	37 (1, 3)	47	4 (1, 3, 4)	43 (1, 3) 6 (1, 4)	72	9 (1, 3, 4)	19 (1, 3)
Xylose	50	24 (1, 3, 4)	3 (1, 3) 12 (1, 4) 10 (2, 3)	24	54 (1, 3, 4)	2 (1, 3) 6 (1, 4) 13 (2, 3)	28	30 (1, 3, 4)	9 (1, 3) 33 (1, 4)	79	10 (1, 3, 4)	4 (1, 3) 6 (1, 4)
Mannose	11	30 (1, 3, 4)	31 (1, 2) 6 (1, 4) 14 (1, 6) 9 (3, 6)	9	29 (1, 3, 4)	25 (1, 2) 6 (1, 4) 3 (1, 6) 25 (3, 6)	20	15 (1, 3, 4)	21 (1, 2)	19	25 (1, 3, 4)	16 (1, 2)
Galactose	33	29 (1, 3, 4)	10 (1, 3) 28 (1, 6)	9	24 (1, 3, 4) 55 (3, 4, 6)	6 (1, 3) 5 (1, 6)	25	36 (1, 3, 4)	29 (1, 3)	55	33 (1, 3, 4)	12 (1, 3)
Glucose	21	64 (1, 3, 4)	15 (1, 3)	18	56 (1, 3, 4)	25 (1, 3)	7	0	80 (1, 3) 14 (1, 2)	39	28 (1, 3, 4)	32 (1, 3)

Each entry shows the percentage per sugar of a particular linkage type, followed (for branched and non-branched types) by numbers in parentheses describing the linkages.

* T, terminal; B, branched; NB, non-branched.

† Furanose form.

‡ Pyranose form.

Previous studies of DOC composition and cycling have emphasized the role of geopolymerization reactions in the production of structurally complex, metabolically resistant organic matter. The recent report of bacterial porin-like proteins dissolved in sea water demonstrates the potential for a contribution from resistant biopolymers¹⁸. Our study extends this concept to a quantitatively significant fraction of DOC, where a family of closely related acyl-oligosaccharides formed by direct biosynthesis contributes up to 70% of UDOM and 20% of total DOC in surface sea water. The very similar monosaccharide composition of surface and deep-sea UDOM suggests that some fraction of these oligosaccharides may survive over long timescales and contribute a significant fraction of the total marine DOC^{12,18}. Further work detailing the chemical structure, potential sources, and metabolism of these oligosaccharides is needed to understand better the link between algal production of DOC, and its accumulation in surface sea water. □

Received 23 July 1996; accepted 26 March 1997.

- Hedges, J. I. Global biogeochemical cycles: progress and problems. *Mar. Chem.* **39**, 67–93 (1992).
- Peltzer, E. & Hayward, N. A. Spatial and temporal variability of total organic carbon along 140° W in the equatorial Pacific Ocean in 1992. *Deep-Sea Res.* **43**, 1155–1180 (1996).
- Chopin-Montegut, G. & Avril, B. Vertical distribution and temporal variation of dissolved organic carbon in the North-Western Mediterranean Sea. *Deep-Sea Res.* **40**, 1963–1972 (1993).
- Williams, P. M. & Gordon, L. I. Carbon-13: carbon-12 ratios in dissolved and particulate organic matter in the sea. *Deep-Sea Res.* **17**, 19–27 (1970).
- Druffel, E. R. M., Williams, P. M., Bauer, J. E. & Ertel, J. R. Cycling of dissolved and particulate organic matter in the open ocean. *J. Geophys. Res.* **97**, 15639–15659 (1992).
- Menzel, D. in *The Sea Vol. 5* (ed. Goldberg, E. D.) 659–678 (Wiley-Interscience, New York, 1974).
- Carlson, C. A., Ducklow, H. W. & Michaels, A. F. Annual flux of dissolved organic carbon from the euphotic zone in the northwestern Sargasso Sea. *Nature* **371**, 405–408 (1994).
- Benner, R., Pakulski, J. D., McCarthy, M., Hedges, J. I. & Hatcher, P. G. Bulk chemical characterization of dissolved organic matter in the ocean. *Science* **255**, 1561–1564 (1992).
- Santschi, P. H. et al. Isotopic evidence for the contemporary origin of high molecular weight organic matter in oceanic environments. *Geochim. Cosmochim. Acta* **59**, 625–631 (1995).
- Buesseler, K. O. et al. An intercomparison of cross-flow filtration techniques used for sampling marine colloids. Overview and organic carbon results. *Mar. Chem.* **55**, 1–32 (1996).
- Sakugawa, H. & Handa, N. Isolation and chemical characterization of dissolved and particulate polysaccharides in Mikawa Bay. *Geochim. Cosmochim. Acta* **49**, 1185–1193 (1985).
- McCarthy, M. D., Hedges, J. I. & Benner, R. The chemical composition of dissolved organic matter in seawater. *Chem. Geol.* **107**, 503–507 (1993).
- Pazur, J. H. in *Carbohydrate Analysis* (eds Chaplin, M. F. & Kennedy, J. F.) 55–96 (IRL, Washington, 1986).
- Painter, T. J. in *The Polysaccharides Vol. 2*, Ch. 4 (ed. Aspinol, G. O.) 196–285 (Academic, New York, 1983).
- Gagosian, R. B. & Stuermer, D. H. The cycling of biogenic compounds and their diagenetically transformed products in seawater. *Mar. Chem.* **5**, 605–632 (1977).
- Harvey, G. R., Boran, D. A., Chesal, L. A. & Tokar, J. M. The structure of marine fulvic and humic acids. *Mar. Chem.* **12**, 119–132 (1983).
- Mopper, K. et al. The role of surface-active carbohydrates in the flocculation of a diatom bloom in a mesocosm. *Deep-Sea Res.* **42**, 47–73 (1995).

- Tanoue, E., Nishiyama, S., Kamo, M. & Tsugita, A. Bacterial membranes: Possible source of a major dissolved protein in seawater. *Geochim. Cosmochim. Acta* **59**, 2643–2648 (1995).
- Vernon-Clark, R. N., Goldberg, E. D. & Bertine, K. K. Organic and inorganic characterization of marine colloids. *Chem. Ecol.* **11**, 69–83 (1996).

Acknowledgements. We thank V. Klap (Netherlands Institute of Ecology) for supplying the Oosterschelde seawater sample, and C. Johnson for assistance with acquiring GC/MS and NMR spectra. This work was supported by the US Department of Energy Ocean Margins Program.

Correspondence and requests for materials should be addressed to D.J.R.

Changes of the Earth's rotation axis owing to advection of mantle density heterogeneities

Bernhard Steinberger* & Richard J. O'Connell

Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts 02138, USA

Polar wander, the secular motion of the Earth's rotation axis relative to its surface, has been studied for many years. Dynamical arguments^{1–3} show that polar wander can arise from the redistribution of mass in a plastic deformable Earth, the rate depending on both the rate of mass redistribution and the rate at which the Earth's rotational bulge can readjust to the changing rotation axis. Here we use a viscosity structure obtained through geoid modelling⁴, a mantle flow field consistent with tomographic anomalies⁵, and time-dependent lithospheric plate motions⁶ to calculate the advection of mantle density heterogeneities and corresponding changes in the degree-two geoid during the Cenozoic era. We show that the rotation axis will follow closely any imposed changes of the axis of maximum non-hydrostatic moment of inertia. The resulting path of the rotation axis agrees well with palaeomagnetic results⁷, with the model predicting a current rate of polar motion that explains 40% of that observed geodetically⁸.

* Present address: Institute of Meteorology and Geophysics, University Frankfurt, 60323 Frankfurt, Germany

The rotational bulge tends to stabilize the location of the pole, but the bulge can readjust by plastic flow to a change in the position of the rotation axis. If the timescale for the adjustment of the bulge is short enough, the location of the rotation pole is determined by the location of the greatest principal moment of inertia based on the non-hydrostatic density distribution (that is, with the contribution of the hydrostatic bulge removed). The non-hydrostatic moments of inertia, however, depend on the density distribution in the Earth's mantle, which may change with time owing to thermal convection and tectonic plate motion. The density distribution in the Earth has been inferred from seismic velocity variations and has been related to the non-hydrostatic geoid with dynamical models^{9,10}. In these models the density anomalies cause flow in the viscous mantle, and a consequence is that the density anomalies are carried by the flow, and their distribution is time-dependent.

The path of the axis of maximum non-hydrostatic moment of inertia in Fig. 1 was calculated using the following steps. (1) The

present-day distribution of shear velocity heterogeneity in the mantle is obtained from seismic tomography⁵. (2) Seismic velocity anomalies are converted to density anomalies^{4,11}. (3) Time-dependent plate velocities and geometries⁶ are used to model plate motions during the past 64 million years. (4) A previously proposed viscosity structure⁴ (Fig. 1) is adopted for the mantle. (5) A mantle flow field that satisfies the boundary conditions on surface plate velocity and is driven by the internal density heterogeneities is calculated¹². (6) The mantle flow field is used to advect the density field back in time from the present. (7) A geoid kernel (Fig. 1) is used to calculate the degree-two non-hydrostatic geoid from the density field at each time of the calculation. The opposite signs of the kernel in the upper and lower mantle mean that areas of low density in the lower mantle and of high density in the upper mantle cause positive geoid anomalies. (8) The difference between the actual and calculated degree-two geoid coefficients is added to obtain agreement of actual and calculated position of the pole at the present time. (9)

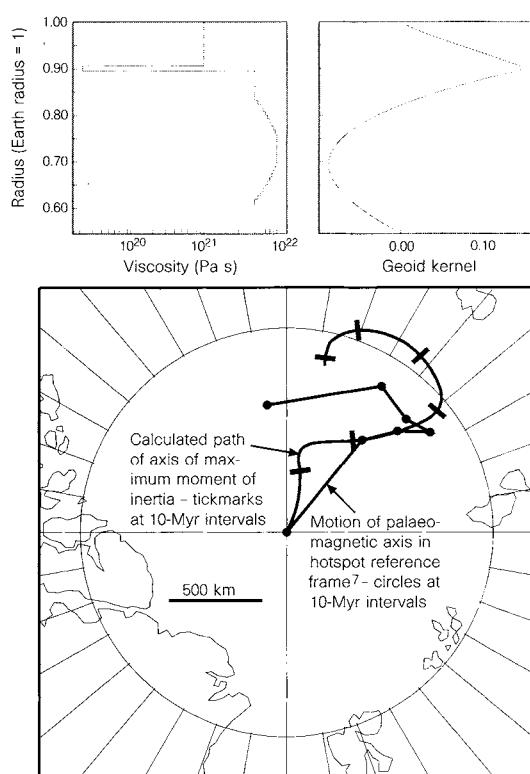


Figure 1 Change of the Earth's rotation axis: results of model calculation compared with palaeomagnetic results⁷. The model is based on flow in a viscous mantle driven by density heterogeneities inferred from the seismic tomography model SH12/WM13 (ref. 5); a conversion factor $(\delta\rho/\rho)/(\delta v/v) = 0.2$ from seismic velocity anomaly (v) to density anomaly (ρ) is used. This value is close to what has been proposed on theoretical grounds and from laboratory experiments¹¹, as well as from modelling the geoid⁴. The flow field is also constrained by realistic plate motions during the Cenozoic era; plate reconstructions⁶ are inferred from sea-floor magnetic anomalies and hotspot traces. The viscosity model used (upper left) is based on models that reconcile the current geoid with tomographic models⁴. The geoid kernel (upper right), which includes the effects of compressibility¹³, shows the relative contribution of density anomalies at a certain depth to the degree-two geoid^{10,13} (which determines the moments of inertia). The calculated polar wander shown depends primarily on the bipolar nature of the geoid kernel as a function of depth. Any viscosity model that produced the same kernel function and had similar maximum viscosity would produce similar results. For other viscosity models considered, the calculated polar motion was too large to be consistent with observations, even if a very high viscosity ($\sim 10^{23}$ Pa s) in the lowermost mantle is assumed. For a larger conversion factor between seismic velocity and density, an increased magnitude of polar motion results, but the direction stays more or less the same. Other tomographic models give similar results, but SH12/WM13 produces the best overall fit. Flow in the upper mantle (largely related to plate motions) and flow in the lower mantle (mostly driven by internal density heterogeneities) produce contributions of similar magnitude to this process.

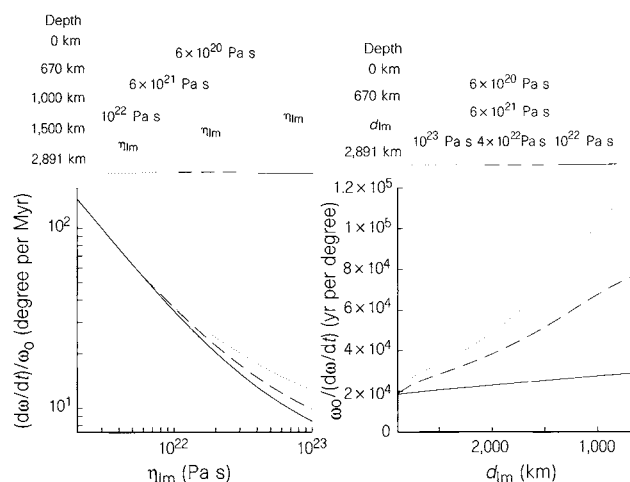


Figure 2 Rate of polar motion for a viscous, incompressible, rotating Earth model that was in hydrostatic equilibrium before a density heterogeneity corresponding to an inertia tensor element $J_{13} = 10^{33}$ kg m² was instantaneously imposed. The rate of polar motion $(d\omega/dt)/\omega_0$ plotted is the relatively uniform rate after initial transients have decayed, and is controlled by the rate at which the hydrostatic bulge can adjust to the changing rotation axis. The diurnal rotation rate is ω_0 , and the model has a mantle of density $4,424$ kg m⁻³ and a core of density $10,989$ kg m⁻³. Left-hand panel, the dependence on lower-mantle viscosity, η_{lm} ; right-hand panel, the influence of the depth, d_{lm} , at which viscosity increases in the lower mantle (this figure plots the inverse of the rate of polar motion). Similar results have been used previously to calculate rates of polar motion excited by random subducted slabs for a variety of Earth models²⁷.

The degree-two non-hydrostatic geoid is converted into the moment of inertia tensor by McCullagh's formula, and the principal axes are calculated.

These calculations are obviously approximate, but they are based on current dynamical models of mantle flow and dynamic support of the geoid^{4,9,10,12,13}. The extension here has been to apply the model to previous times and plate configurations by advecting the density field backwards in time. This procedure should be accurate for relatively short times, although it will obviously break down for times longer than it would take a thermal anomaly to develop and rise or sink through most of the mantle. In a similar approach, subduction history has been used to calculate polar motion over an even longer time interval¹⁴.

The rotation axis will follow the axis of the largest moment of inertia on a timescale governed by the readjustment of the rotational bulge. This in turn depends primarily on the viscosity structure of the mantle³. The following considerations most probably rule out a significant difference between rotation axis and axis of maximum moment of inertia.

The rotation axis tends to be nearly aligned with the axis of maximum total moment of inertia^{2,3}, which is the sum of the non-hydrostatic moment of inertia and the larger hydrostatic moment determined by the rotational bulge. The non-hydrostatic moment of inertia consists of an imposed part (which we calculated from the density and flow models) and a part that is due to the delayed viscoelastic relaxation of the bulge after a change of the rotation axis. As the axis of the imposed part of the inertia tensor moves, the rotation axis lags behind it by the angle¹⁵

$$\theta \approx \frac{d\varphi}{dt} \cdot \frac{1}{\Delta J_{\max} \cdot c_{\text{tpw}}} \quad (1)$$

Here $d\varphi/dt$ is the rate at which the axis of the maximum imposed non-hydrostatic moment of inertia moves, $\Delta J_{\max}$ is the difference between maximum and minimum principal moments of inertia from the imposed mass anomalies, and the constant c_{tpw} describes the rate at which the bulge adjusts and hence the rate at which the true polar wander occurs in response to a change in the inertia tensor. It can be calculated for Earth models with various viscosity structures. If viscoelastic relaxation can be expressed in terms of eigenmodes, c_{tpw} can be expressed by

$$c_{\text{tpw}}^{-1} = \sum_j \frac{c_{j0}}{s_j} J_j \quad (2)$$

where c_{j0} are the eigenmode expansion coefficients of the equilibrium shape minus the elastic deformation that would result if the Earth instantaneously started rotating at its present rate, s_j are decay constants and J_j are differences between maximum and minimum principal moments of inertia for these eigenmodes. Otherwise c_{tpw} may be determined numerically.

If a mass anomaly is emplaced suddenly, and the rotation vector is primarily in the z -direction, its rate of change will be given by

$$\frac{d\omega_x}{dt} = c_{\text{tpw}} \omega_0 J_{13}^0 \quad (3)$$

where ω_0 is the diurnal rotation rate and J_{13}^0 is the non-diagonal inertia tensor component due to the emplaced non-hydrostatic excess masses. This equation applies to the quasi-steady state motion of the pole after initial transients have died out (that is, when the buildup of the non-hydrostatic shape due to the change in the rotation axis is exactly compensated for by viscoelastic decay of the old bulge).

Figure 2 shows the steady state rate of polar motion ($d\omega_x/dt \cdot (1/\omega_0)$) caused by an imposed mass for a variety of Earth models. The rate (and hence c_{tpw}) is inversely proportional to the viscosity of the lower mantle, roughly inversely proportional to the thickness of a high-viscosity layer in the lower mantle and only slightly dependent on upper-mantle viscosity structure.

The Earth's rotation axis moved less than $\sim 10^\circ$ in the hotspot reference frame^{7,16,17} during the past 50 Myr; equation (1) and inferred values of c_{tpw} from Fig. 2 show that the rotation pole would lag the inertia axis by $\theta < 1^\circ$, even for models with $\eta \approx 10^{23}$ Pa s in the lower mantle.

It should be noted that equation (3) applies only if the emplacement of mass anomalies is rapid compared with the change of the rotation axis, as would be the case for a relatively high mantle viscosity. We can therefore estimate an upper bound to mantle viscosity. For the observed motion of $\sim 10^\circ$ in 50 Myr (refs 7, 16), the cases considered in Fig. 2, left, yield an upper bound to lower-mantle viscosity between about 3 and 5×10^{25} Pa s. This is not a stringent bound, as considerably slower motion of the rotation axis cannot be excluded¹⁷, but it is similar to values obtained before³.

If the mass anomalies are emplaced continuously at a rate corresponding to a growth rate of the inertia tensor component J_{13} , the maximum rate of polar motion

$$\left. \frac{d\omega}{dt} \right|_{\max} \approx 0.47 \cdot \sqrt{J_{13} c_{\text{tpw}}} \quad (4)$$

occurs when the two larger principal non-hydrostatic moments of inertia are almost equal. The difference between instantaneous and gradual emplacement is illustrated in the bottom part of Fig. 3.

Figure 3 shows viscoelastic relaxation and corresponding changes in the rotation axis for a number of cases. We find that the speed at which the rotation axis moves is mainly determined by lower-mantle viscosity, whereas the effects of introducing an internal boundary or varying upper-mantle viscosity, and the effects of differences between viscous and viscoelastic Earth models, are all comparatively small.

The steady-state approximation is valid for times longer than the largest decay time. Unless internal modes are present (for example where there is a chemical boundary; Fig. 3) the largest decay time is small compared with timescales of polar wander, and steady-state is a good approximation. Furthermore, the magnitude of the effect of a chemical boundary is relatively small. We therefore conclude that for a realistic Earth model the angle θ is most likely to be small, even if polar motion is somewhat slowed by an internal chemical boundary^{18,19}, and the path of the axis of maximum imposed moment of inertia can be compared directly with palaeomagnetic results⁷.

Observations of polar wander have been inferred from palaeomagnetic apparent polar wander paths, assuming that the rotation and palaeomagnetic axes are essentially the same²⁰. If there were no plate motions, the interpretation of these would be straightforward, but the determination of polar wander on a planet with a mobile surface is more complicated. Two reference frames have been proposed for plate motions: a mean lithospheric frame, which is the average of all plate motions, and a hotspot frame that assumes that hotspots are fixed in location by a lower mantle that is essentially stationary relative to plate motions. Early work suggested that there is no significant motion of the rotation axis in the mean lithospheric frame^{21,22}, whereas there is motion of the rotation axis with respect to the hotspot frame²³. More recent results, however, indicate that both reference frames and the rotation axis all move with respect to each other²⁴. Results from models of plumes and hotspots indicate similar motions¹⁵.

The comparison between the model results and palaeomagnetic observations of polar wander is shown in Fig. 1. The scatter between the palaeomagnetic curves from individual continents is larger than the average motion¹⁶, and there are also substantial differences between results by various authors^{7,16,17}. The good agreement between modelled and observed polar motion in Fig. 1 was obtained with a geoid kernel that reverses sign, with roughly equal magnitude on either side, as shown in Fig. 1; this behaviour is a consequence of the viscosity model used, which has also been used to model the present geoid⁴.

The rate of polar motion in recent times has been directly

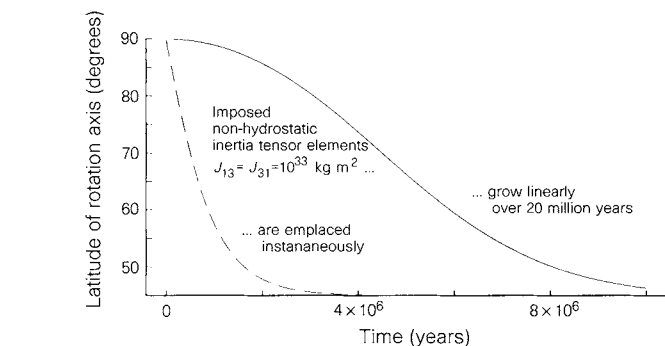
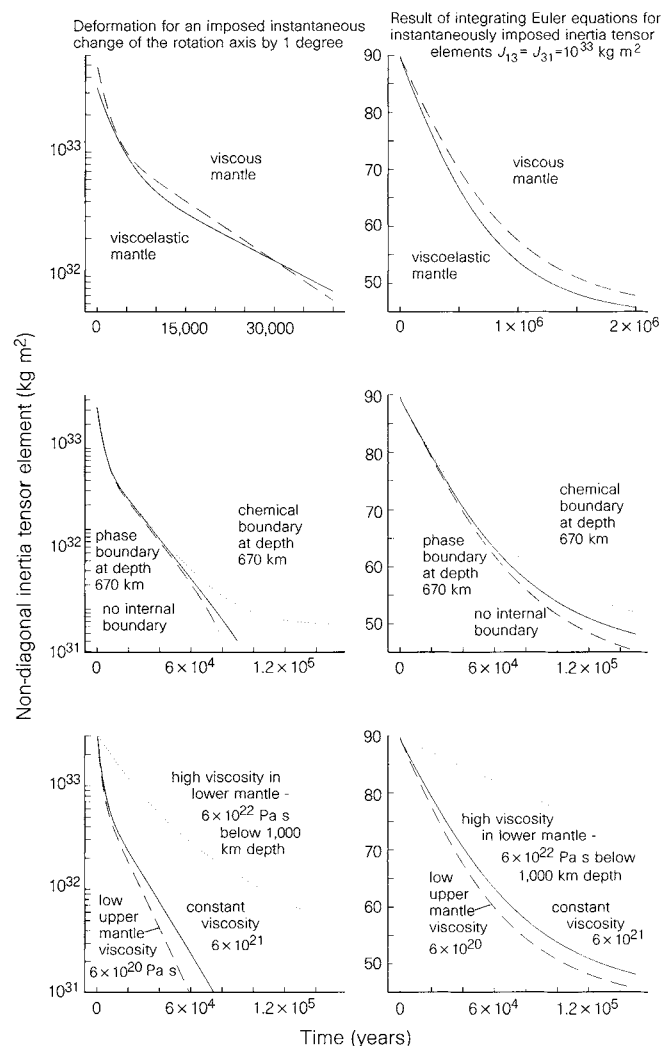


Figure 3 The effects of various model parameters on viscoelastic relaxation and the motion of the rotation axis. Left, the non-diagonal inertia tensor components as a function of time following a 1° shift of the rotation axis. Right, the motion of the rotation pole following the emplacement of density heterogeneities that ultimately drive the pole to move 45°. Top row, a viscous incompressible mantle (as in Fig. 2) is compared with a viscoelastic mantle parameterized as a PREM (ref. 28) lower mantle. Viscosity is 10^{21} Pa s. Because of the instantaneous elastic deformation, there is an offset of the two curves at time zero in the left plot, and the rotation axis moves faster in the viscoelastic case. Second row, the effects of an internal boundary at 670 km depth. The density parameters above the boundary are as in PREM between 5,701 and 5,771 km. Shown are the effects of an equilibrium phase boundary and a chemical boundary with a density contrast. This latter models shown a very slow relaxation mode. The calculated polar motion is initially similar in all cases, but the curve for a chemical boundary layer flattens out faster, as steady state is approached, corresponding to previous results that find polar motion to be significantly slower for a chemical than for a phase boundary^{18,19}. Third row, the effects of various viscosity models; all other parameters similar to the viscoelastic case in the top row. Fourth row, the time response for instantaneous emplacement of density heterogeneity compared with the response when the heterogeneity grows linearly over 20 Myr. These results were obtained using a time domain approach^{15,29} for a mantle with a Maxwell viscoelastic rheology.

measured by astronomical and geodetic techniques⁸, and it is roughly in accord with the longer-term motion inferred from palaeomagnetic observations²⁵. The current polar motion is an important constraint on current models of mantle rheology inferred from observations of post-glacial isostatic readjustment^{26,30}. Our results yield a current rate of polar motion of 0.39° per Myr towards 24W, with an average of 0.37° per Myr towards 24W over the past 1 Myr. This is a substantial fraction of the value 0.95° per Myr towards 76W that is used to constrain rebound models²⁶. Figure 1 shows that rates of this magnitude characterize the polar motion inferred from palaeomagnetic results as well as the model results for the past 50 Myr or so. Thus a considerable fraction of the current polar motion may represent a secular trend that has existed for millions of years. The effects of post-glacial rebound would be superposed on the long-term trend. Coincidentally, the two motions are similar²⁵. Clearly both effects need to be well understood to interpret the current observations properly.

It is apparent that a plausible model of mantle viscosity, coupled with the time-dependent density distribution implied by the model, can account for long-term polar wander similar to that observed. Similar models can predict polar motion that is considerably faster than observed as well. The remarkable thing about true polar wander may not be that it occurs, but that it is so slow. Our current models of the Earth's mantle, however, allow us to predict the polar wander path reasonably well. □

Received 31 October 1996; accepted 18 March 1997.

- Gold, T. Instability of the Earth's axis of rotation. *Nature* **175**, 526–529 (1955).
- Munk, W. & MacDonald, G. J. F. *The Rotation of the Earth* (Cambridge Univ. Press, New York, 1960).

- Goldreich, P. & Toomre, A. Some remarks on polar wandering. *J. Geophys. Res.* **74**, 2555–2567 (1969).
- Forte, A. M., Dziewonski, A. M. & Woodward, R. L. Aspherical structure of the mantle, tectonic plate motions, nonhydrostatic geoid and topography of the core–mantle boundary in *Dynamics of the Earth's Deep Interior and Earth Rotation*, *Geophys. Mon. Ser.* Vol. 72, (eds Le Mouél, J.-L. et al.) 135–166 (AGU, Washington DC, 1994).
- Su, W., Woodward, R. L. & Dziewonski, A. M. Degree 12 model of shear velocity heterogeneity in the mantle. *J. Geophys. Res.* **99**, 6945–6980 (1994).
- Gordon, R. G. & Jurdy, D. M. Cenozoic global plate motions. *J. Geophys. Res.* **91**, 12389–12406 (1986).
- Besse, G. & Courtillot, V. Revised and synthetic apparent polar wander paths of the African, Eurasian, North American and Indian plates, and true polar wander since 200 Ma. *J. Geophys. Res.* **96**, 4029–4050 (1991).
- Dickman, S. R. Secular trends in the Earth's rotation pole: Consideration of the motion of latitude observatories. *Geophys. J. R. Astron. Soc.* **51**, 229–224 (1977).
- Hager, B. H., Clayton, R. W., Richards, M. A., Comer, R. P. & Dziewonski, A. M. Lower mantle heterogeneity, dynamic topography and the geoid. *Nature* **313**, 541–545 (1985).
- Richards, M. & Hager, B. Geoid anomalies in a dynamic Earth. *J. Geophys. Res.* **89**, 5987–6002 (1984).
- Karato, S. Importance of anelasticity in the interpretation of seismic tomography. *Geophys. Res. Lett.* **20**, 1623–1626 (1993).
- Hager, B. H. & O'Connell, R. J. A simple global model of plate dynamics and mantle convection. *J. Geophys. Res.* **86**, 4843–4867 (1981).
- Panasuk, S. V., Hager, B. H. & Forte, A. M. Understanding the effects of mantle compressibility on geoid kernels. *Geophys. J. Int.* **124**, 121–133 (1996).
- Richards, M. A., Ricard, Y., Lithgow-Bertelloni, C., Spada, G. & Sabadini, R. An explanation for Earth's long-term rotational stability. *Science* **275**, 372–375 (1997).
- Steinberger, B. M. Motion of hotspots and changes of the Earth's rotation axis caused by a convecting mantle. Thesis, Univ. Harvard (1996).
- Andrews, J. A. True polar wander: An analysis of cenozoic and mesozoic paleomagnetic poles. *J. Geophys. Res.* **90**, 7737–7750 (1985).
- Livermore, R. A., Vine, F. J. & Smith, A. G. Plate motions and the geomagnetic field, II, Jurassic to Tertiary. *Geophys. J. R. Astron. Soc.* **79**, 939–961 (1984).
- Ricard, Y. & Sabadini, R. Rotational instabilities of the Earth induced by mantle density anomalies. *Geophys. Res. Lett.* **17**, 627–630 (1990).
- Sabadini, R. & Yuen, D. A. Mantle stratification and long-term polar wander. *Nature* **339**, 373–375 (1989).
- Tarling, D. H. *Principles and Applications of Paleomagnetism* (Chapman and Hall, London, 1971).
- McElhinny, M. W. Mantle plumes, palaeomagnetism and polar wandering. *Nature* **241**, 523–528 (1973).

22. Jurdy, D. M. & Van der Voo, R. A method for the separation of true polar wander and continental drift including results for the last 55 m.y. *J. Geophys. Res.* **79**, 2945–2952 (1974).
23. Duncan, R. A., Petersen, N. & Hargraves, R. B. Mantle plumes, movement of the European plate and polar wandering. *Nature* **239**, 82–86 (1972).
24. Gordon, R. G. & Livermore, R. A. apparent polar wander of the mean-lithosphere reference frame. *Geophys. J. R. Astron. Soc.* **91**, 1049–1057 (1987).
25. Gordon, R. G. Plate motions, crustal and lithospheric mobility, and paleomagnetism: Prospective viewpoint. *J. Geophys. Res.* **100**, 24367–24392 (1995).
26. Peltier, W. R. & Jiang, X. Glacial isostatic adjustment and Earth rotation: Refined constraints on the viscosity of the deepest mantle. *J. Geophys. Res.* **101**, 3269–3290 (1996).
27. Spada, G., Ricard, Y. & Sabadini, R. Excitation of true polar wander by subduction. *Nature* **360**, 452–454 (1992).
28. Dziewonski, A. M. & Anderson, D. L. Preliminary Reference Earth Model. *Phys. Earth Planet. Inter.* **25**, 297–356 (1981).
29. Hanyk, L., Moser, J., Yuen, D. A. & Matyska, C. Time-domain approach for the transient responses in stratified viscoelastic Earth. *Geophys. Res. Lett.* **22**, 1285–1288 (1995).
30. Mitrova, J. X. & Forte, A. M. Radial profile of mantle viscosity: Results from the joint inversion of convection and post-glacial rebound observables. *J. Geophys. Res.* **102**, 2751–2769 (1997).

Acknowledgements. We thank S. V. Panasyuk for calculating the geoid kernels for us, and J. X. Mitrova for a review. This work was supported by the NSF.

Correspondence should be addressed to B.S. (e-mail: steinber@geophysik.uni-frankfurt.de).

Body mass and encephalization in Pleistocene *Homo*

Christopher B. Ruff*, Erik Trinkaus† & Trenton W. Holliday‡

* Department of Cell Biology and Anatomy, Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205, USA

† Department of Anthropology, University of New Mexico, Albuquerque, New Mexico 87131, USA, and URA 376 du C.N.R.S., Université de Bordeaux I, 33405 Talence, France

‡ Department of Anthropology, College of William and Mary, Williamsburg, Virginia 23187-8795, USA

Many dramatic changes in morphology within the genus *Homo* have occurred over the past 2 million years or more, including large increases in absolute brain size and decreases in postcanine dental size and skeletal robusticity. Body mass, as the 'size' variable against which other morphological features are usually judged, has been important for assessing these changes^{1–5}. Yet past body mass estimates for Pleistocene *Homo* have varied greatly, sometimes by as much as 50% for the same individuals^{2,3,6–12}. Here we show that two independent methods of body-mass estimation yield concordant results when applied to Pleistocene *Homo* specimens. On the basis of an analysis of 163 individuals, body mass in Pleistocene *Homo* averaged significantly (about 10%) larger than a representative sample of living humans. Relative to body mass, brain mass in late archaic *H. sapiens* (Neanderthals) was slightly smaller than in early 'anatomically modern' humans, but the major increase in encephalization within *Homo* occurred earlier during the Middle Pleistocene (600–150 thousand years before present (kyr BP)), preceded by a long period of stasis extending through the Early Pleistocene (1,800 kyr BP).

It is generally acknowledged, even by those who have used other methods, that the best means of estimating body mass from skeletal or fossil remains, when feasible, is to use features that have some direct functional relationship to body mass^{9,11,12}. For hominids, the skeletal dimensions used most often have been lower limb long bone diaphyseal and articular breadths^{2,3,6,7,9}. Diaphyseal breadths of fossil hominids are problematic as body mass estimators because relative to body size they are systematically larger than modern humans, probably as a response to increased mechanical loading⁵. In contrast, articulations are much less environmentally sensitive^{13,14}, and thus are potentially better body-size indicators. The articular dimension used here as a body-mass estimator is femoral head

breadth because it is available for many fossil *Homo* specimens, is easily measured and highly reproducible, and because several investigators have provided information on the relationship between femoral head breadth and body mass in modern humans^{6,13,15} (see Methods).

The second method used here to estimate body mass does not rely on any assumptions about the mechanical relationship between a particular skeletal feature and body size (support of body weight). Rather, in this approach body mass is estimated directly from reconstructed stature and body breadth. A modern worldwide anthropometric sampling of 56 population/sex-specific means¹⁶ was used to derive multiple regressions of body mass on stature and bi-iliac (maximum pelvic) breadth (Methods).

Figure 1 compares femoral head and stature/bi-iliac estimates of body mass for 75 Pleistocene *Homo* specimens. The mean absolute difference between estimates is about 5 kg (7.6%), and the mean directional difference is less than 1 kg (1.1%). Paired *t*-tests between results are not significant ($P \geq 0.30$). Thus, equations based on femoral head size and stature/bi-iliac breadth yield similar body mass estimates when applied to Pleistocene *Homo*, with very little systematic bias. Because the two techniques are based on different rationales and skeletal dimensions, yet nevertheless converge on the same result, this increases confidence in both.

Skeletal dimensions for 163 Pleistocene *Homo* specimens, dated 10–1,950 kyr BP, were derived from previously published sources and personal measurements^{5,10,17}. Most regions of the Old World (except Australia) are represented, although the majority of the sample is from Europe (55%), with the remainder from Africa (27%), western Asia (15%) and eastern Asia (3%). (Data for individual specimens are given in Supplementary Information.) The resulting body mass estimates are shown in Fig. 2a, together with 51 sex/population-specific means for a worldwide sampling of living humans (ref. 16, excluding five Pygmy data points). More than three-quarters (125/163) of the Pleistocene specimens fall above the living human mean. On average, Pleistocene specimens are 7.4 kg larger (mean $\pm$ s.e., 65.6 ± 0.7 kg) than living humans (58.2 ± 1.0 kg), a highly significant 12.7% difference in body mass ($P < 0.0001$, *t*-test).

There is some indication in Fig. 2a that body mass is lower in the Early Pleistocene and rises to peak values in the Late Pleistocene. However, this is largely an artefact of two confounding variables: sex

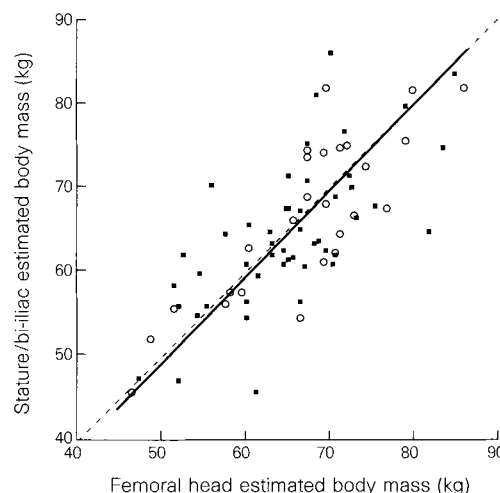


Figure 1 Comparison between body mass estimated from stature and bi-iliac breadth, and body mass estimated from femoral head breadth in 75 Pleistocene *Homo* specimens. Empty symbols, measured bi-iliac breadth; filled symbols, estimated bi-iliac breadth. Solid line, reduced major axis regression ($y = 1.04 \times x - 3.4$; $r = 0.738$). Dotted line indicates equivalence of y and x .

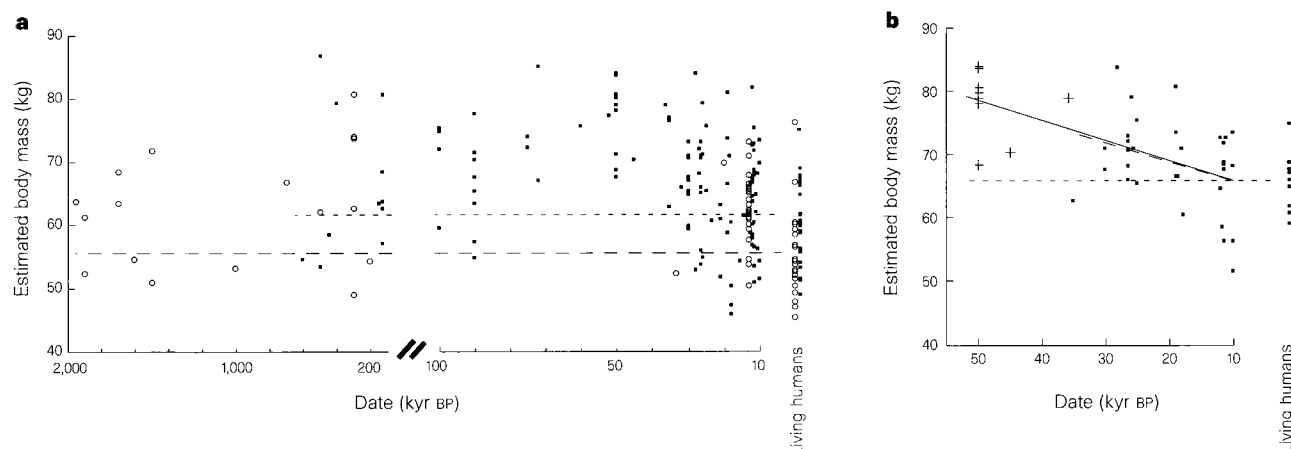


Figure 2 Changes through time in body mass of *Homo*. **a**, Total sample. Empty symbols, lower latitude ($<30^{\circ}\text{N}$) (dashed line through living mean); closed symbols, higher latitude (dotted line through living mean). Note change in temporal (x) axis scale at 100 kyr BP. **b**, Higher-latitude males. Crosses, archaic

H. sapiens; squares, early anatomically modern (EAM) and living *H. sapiens* (dotted line through living mean). Solid line, least-squares regression through total Pleistocene sample ($y = 0.318 \times x + 62.6$; $r = 0.582$); dashed line, least-squares regression through EAM sample ($y = 0.299 \times x + 62.9$; $r = 0.334$).

Table 1 Body-mass and brain-mass data

Sample	Temporal range (kyr BP)	Body mass (kg)* (mean $\pm$ s.e. (n))	Cranial capacity (cc)† (mean $\pm$ s.e. (n))	Brain mass (g)‡ (mean $\pm$ s.e. (n))	Encephalization quotient‡	
					Associated specimens (mean $\pm$ s.e. (n))	Sample means (mean)
Living worldwide	—	58.2 $\pm$ 1.0 (51)	1,349	1,302	—	5.288
Pecos Pueblo	—	55.5 $\pm$ 1.2 (29)	1,308 $\pm$ 23 (29)	1,263 $\pm$ 22 (29)	5.349 $\pm$.103 (29)	—
Late Upper Paleolithic	10–21	62.9 $\pm$.9 (71)	1,466 $\pm$ 35 (23)	1,412 $\pm$ 33 (23)	5.479 $\pm$.083 (18)	5.406
Early Upper Paleolithic	21–35	66.6 $\pm$ 1.3 (33)	1,517 $\pm$ 30 (15)	1,460 $\pm$ 28 (15)	5.467 $\pm$.142 (10)	5.352
Late archaic <i>H. sapiens</i>	36–75	76.0 $\pm$ 1.4 (17)	1,498 $\pm$ 45 (14)	1,442 $\pm$ 42 (14)	4.984 $\pm$.165 (8)	4.781
Skhul-Qafzeh	90	66.6 $\pm$ 2.2 (10)	1,501 $\pm$ 45 (6)	1,444 $\pm$ 42 (6)	5.369 $\pm$.083 (4)	5.293
early Late Pleistocene	100–150	67.7 $\pm$ 2.4 (10)	1,354 $\pm$ 41 (8)	1,307 $\pm$ 39 (8)	4.682 (1)	4.732
late Middle Pleistocene	200–300	65.6 $\pm$ 5.1 (6)	1,186 $\pm$ 32 (17)	1,148 $\pm$ 30 (17)	—	4.257
middle Middle Pleistocene	400–550	67.9 $\pm$ 6.4 (5)	1,090 $\pm$ 38 (12)	1,057 $\pm$ 36 (12)	—	3.818
late Early to early Middle Pleistocene	600–1,150	58.0 $\pm$ 4.3 (3)	856 $\pm$ 52 (7)	835 $\pm$ 50 (7)	—	3.400
Early Pleistocene	1,200–1,800	61.8 $\pm$ 4.0 (5)	914 $\pm$ 45 (5)	890 $\pm$ 43 (5)	3.064 (1)	3.458

* Data from Fig. 2, except that Early Pleistocene sample does not include earliest (pre-1,800 kyr BP) specimens (see text). Pecos Pueblo body weights from femoral head and bi-iliac/stature formulae.

† Average cranial capacity for living worldwide sample from Beals *et al.*²⁸; Pecos Pueblo cranial capacities from Hooton's original grave cards³⁰; Pleistocene *Homo* cranial capacities from the literature and personal measurement.

‡ See Methods for derivation of brain mass from cranial capacity, and encephalization quotient.

and geography. Modern and fossil *Homo* males are larger than females, and there is a bias towards males in our Late Pleistocene sample (72 of 114 sexable specimens). Ecogeographic patterning in body mass has been demonstrated in modern humans, with populations from higher latitudes being larger on average than those from lower latitudes¹⁶. The same is true for Pleistocene *Homo*: in our sample specimens from over 30°N latitude are significantly larger than those from under 30°N latitude (ANCOVA, controlling for sex and date, $P < 0.005$). All of the Pleistocene *Homo* specimens included here and dated before 600 kyr BP are from tropical regions (Fig. 2a). The apparent sudden increase in average body mass during the later Middle Pleistocene is largely a result of the inclusion of higher-latitude specimens beginning during this time period. If the effects of latitude and sex are controlled through ANCOVA, the pooled Pleistocene *Homo* sample is still 9.2% larger on average than living humans ($P < 0.0001$).

Body mass appears to decline after about 50 kyr BP. Figure 2b shows that within higher-latitude ($>30^{\circ}\text{N}$) males there is a significant decrease (16%, $P < 0.0001$) in average body mass between 50 and 10 kyr BP. The trend is very similar whether or not archaic *Homo sapiens* (Neanderthals) are included. By the Terminal Pleistocene (10–15 kyr BP) body mass is not significantly different from that of living higher-latitude males (0.3% difference). The same is also true of higher-latitude females (3.7% difference). However, lower-

latitude Terminal Pleistocene males and females are still significantly larger (11–12%, $P < 0.01$) than lower-latitude living samples.

The body-mass data are paired with cranial capacity and brain-mass data in Table 1, temporally subdivided within the Pleistocene. Because of uncertainty regarding the cranial–postcranial affinities of very early *Homo*¹⁸, only specimens dated to 1,800 kyr BP or later are included in this analysis. (All individual cranial data are listed in Supplementary Information.)

Early to early Middle Pleistocene (1,800–600 kyr BP) *Homo* was about one-third less encephalized than Recent humans, and there was no increase in encephalization quotient (EQ) throughout this time period (Table 1). By the early Late Pleistocene (150–100 kyr BP), EQ had increased to values within about 10% of those of Recent humans. However, early Late Pleistocene and late archaic *H. sapiens* are significantly lower in EQ than pooled 'early anatomically modern' (EAM) *H. sapiens* (Skhul-Qafzeh and Upper Palaeolithic samples) ($P < 0.01$, *t*-test of individually associated specimens). The same results holds true whether or not Recent humans are pooled with EAM *H. sapiens*, or the early Late Pleistocene individual with associated data (Tabun C1) is pooled with late archaic *H. sapiens*.

Because EQs in general, and any EQ in particular, are subject to a number of limitations, encephalization within *Homo* was also evaluated by plotting $\log(\text{cranial capacity})$ against $\log(\text{body mass})$

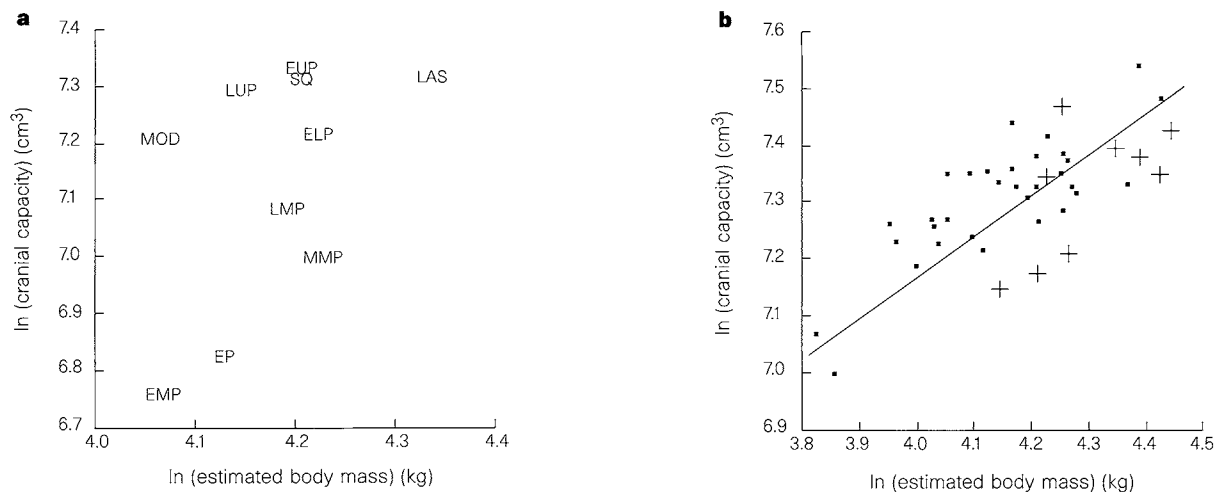


Figure 3 Log-transformed brain mass versus body mass in Pleistocene *Homo*. **a**, Temporal group means. MOD, living humans; LUP, Late Upper Palaeolithic; EUP, Early Upper Palaeolithic; SQ, Skhul-Qafzeh; LAS, late archaic *H. sapiens*; ELP, early Late Pleistocene; LMP, late Middle Pleistocene; MMP, middle Middle

Pleistocene; EMP, early Middle Pleistocene; EP, early Pleistocene (see Table 1). **b**, Late Pleistocene individuals. Squares, pooled EUP, LUP and SQ individuals; crosses, pooled LAS and ELP individuals; solid line, reduced major axis regression through total pooled sample.

(Fig. 3). Figure 3a is a plot of the sample mean data from Table 1. The data appear to be best characterized by three major 'trajectories': an Early to middle Middle Pleistocene trajectory, a late Middle to early Late Pleistocene to late archaic *H. sapiens* trajectory, and a trajectory that includes Recent and EAM *H. sapiens*. Variation in cranial capacity between these three groups with body mass as a covariate is highly significant ($P < 0.0001$, ANCOVA; Tukey tests, $P < 0.01$, all pairwise comparisons). Alternatively, the middle and late Middle Pleistocene groups could be viewed as transitional between an earlier (EP–EMP) and later (ELP–LAS) trajectory, with Recent and EAM *H. sapiens* again forming a third trajectory; this interpretation would be more consistent with the EQ data in Table 1. Figure 3b is a plot of all of the individually associated brain and body masses for the Late Pleistocene samples. Although there is substantial overlap between archaic and EAM *H. sapiens*, they are significantly different ($P < 0.01$); this result holds true whether or not the one early Late Pleistocene specimen is included.

Pilbeam and Gould¹ suggested that the scaling of brain size to body size in *Homo* followed a single log-linear trajectory that eventually separated humans from australopithecines as well as other primates. Walker¹⁹, in part on the basis of the discovery of the early *Homo* KNM-WT 15000, proposed that *Homo* was characterized by two brain-size/body-size scaling trajectories or 'grades': one for *H. habilis*/*H. erectus* and one for modern *H. sapiens*. With more specimens and more accurate means of estimating body mass, it now appears that there were at least three trajectories, or grades of brain size relative to body size within *Homo*, even excluding very early *H. habilis* or *H. rudolfensis* specimens, which may represent another grade⁴. Other previously proposed temporal trends may also now be evaluated with more confidence. For example, the presence of a directional trend in absolute brain size in Early to Middle Pleistocene *Homo* has been debated, with interpretations hampered by the lack of body-mass data for the same time period^{20,21}. Our results support a stasis in relative brain size within *Homo* between 1,800 and at least 600 kyr BP. Opinions regarding the relative encephalization of Late Pleistocene archaic *H. sapiens* (Neanderthals) have varied widely^{9,12}. Our findings indicate that Neanderthals were slightly less encephalized than Recent and EAM *H. sapiens*, but closer in this respect to modern humans than to middle Middle Pleistocene and earlier *Homo*.

Our results also indicate that a decrease in average absolute brain size over the past 35,000 years within *H. sapiens* was paralleled by a

corresponding decrease in average body size, supporting earlier suggestions of a general correlated size reduction in the human skeleton since the early Upper Palaeolithic²². This decrease continued through the Neolithic, at least in Europe²³. Recent secular increases in body size have characterized European and many other higher-latitude populations, whereas many tropical populations have experienced flat or even negative secular trends in size over the same time period²⁴. Our body mass results for tropical and higher latitude samples are consistent with these recent trends. Viewed in this light, although some living humans may be expressing a genetic potential for greater body size retained from earlier hunter-gatherer (Upper Palaeolithic) ancestors²⁵, this phenomenon has been largely limited to higher-latitude populations. □

Methods

Femoral head estimation of body mass. Equations from three studies of a diverse sampling of modern humans were used^{6,13,15}. In one study, estimates had been derived for males and females separately¹³; in the other two only a combined sex sample had been analysed^{6,15}. For the Pleistocene sample, the two combined sex equations were always used and the male or female equation used when sex could be determined (otherwise, a mean of the male and female equation results was used). All equations are for raw (non-logged) data (BM, body mass; FH, femoral head breadth (mm)):

$$BM = 2.239 \times FH - 39.9; r = 0.98 \text{ (derived by us from raw published data) (ref. 6)} \quad (1)$$

$$BM = 2.268 \times FH - 36.5; r = 0.92 \text{ (ref. 15)} \quad (2)$$

$$BM = 2.741 \times FH - 54.9; r = 0.50 \text{ (males);} \quad (3)$$

$$BM = 2.426 \times FH - 35.1; r = 0.411 \text{ (females) (ref. 13)}$$

Estimates from equation (3) were adjusted downwards by 10% as recommended by the authors. When applied to 94 Pleistocene *Homo* specimens with intact femoral heads, correlations between pairs of body mass estimates using the above equations were all 0.97 or better, with a mean difference between results of 4%. The mean of the three estimates was used in the study.

Stature/bi-iliac estimation of body mass. The stature/bi-iliac equations were derived from modern anthropometric data given in ref. 16 (BM, body mass (kg); ST, stature (cm); BI, bi-iliac breadth (cm)):

$$BM = 0.373 \times ST + 3.033 \times BI - 82.5; r = 0.90 \text{ (males);} \quad (4)$$

$$BM = 0.522 \times ST + 1.809 \times BI - 75.5; r = 0.82 \text{ (females)}$$

(Note that the equation for females is different from that published in ref. 16 because of correction of an error in one of the original data points (Aleut

females: correct body mass = 53.4 kg)). The sex-specific formula was used when possible; otherwise the mean of the male and female formulae was used. Before applying these formulae, skeletal bi-iliac breadth was converted to living bi-iliac breadth using the equation (both dimensions in cm):

$$\text{living BI} = 1.17 \times \text{skeletal BI} - 3 \quad (5)$$

derived from comparisons within modern humans¹⁶. Bi-iliac breadth of the Pleistocene specimens was either measured directly or estimated from closely related fossils and/or from known clinal variation in bi-iliac breadth^{16,17}. Stature was estimated from preserved long bone lengths using equations derived from appropriately proportioned modern reference samples. Details are given in Supplementary Information.

Combined body mass estimates. When both an intact femoral head and bi-iliac breadth were available, the mean of the femoral head and stature/bi-iliac estimates was used ($n = 26$). Otherwise, either the femoral head, when available ($n = 67$) or the estimated stature/bi-iliac ($n = 70$) estimate was used.

Brain mass and EQ. Brain mass was derived from cranial capacity using a least-squares regression of 27 primate species that had data available for both parameters^{26,27}, corrected for logarithmic transformation bias:

$$\text{brain mass} = 1.147 \times \text{cranial capacity}^{0.976} \quad (r^2 = 0.995) \quad (6)$$

Encephalization quotient was derived from Martin's²⁸ relationship between brain mass (g) and body mass (kg) in mammals:

$$\text{EQ} = \text{brain mass} / (11.22 \times \text{body mass}^{0.76}) \quad (7)$$

Encephalization quotients (EQ) relating brain mass to body mass were derived in two ways, using individually associated crania and postcrania, and using mean brain mass and body mass within temporally defined groups. The EQs derived using the group means are based on many more specimens, but because they do not use individually matched data they could potentially be biased in several ways. However, other methods of limiting these samples, for example by using only individuals from the same sites and/or sex, produce similar results. Also, where they can be compared, the individually associated and mean data values for the same temporal periods are similar (Table 1). For associated specimens, sex and latitudinal biases in EQ should be minimal: within the Pecos and Pleistocene EAM samples sex differences in EQ average less than 2%, and among modern humans higher- and lower-latitude populations appear to average less than a 4% difference in EQ (mean data from Beals *et al.*²⁹ and our worldwide sample¹⁶).

Received 8 October 1996; accepted 12 March 1997.

- Pilbeam, D. & Gould, S. J. Size and scaling in human evolution. *Science* **186**, 892–901 (1974).
- McHenry, H. M. Early hominid body weight and encephalization. *Am. J. Phys. Anthropol.* **45**, 77–84 (1976).
- McHenry, H. M. In *Evolutionary History of the "Robust" Australopithecines* (ed. Grine, F. E.) 133–148 (Aldine de Gruyter, New York, 1988).
- McHenry, H. Behavioral ecological implications of early hominid body size. *J. Hum. Evol.* **27**, 77–87 (1994).
- Ruff, C. B., Trinkaus, E., Walker, A. & Larsen, C. S. Postcranial robusticity in *Homo*, I: temporal trends and mechanical interpretation. *Am. J. Phys. Anthropol.* **91**, 21–53 (1993).
- McHenry, H. M. Body size and proportions in early hominids. *Am. J. Phys. Anthropol.* **87**, 407–431 (1992).
- Rightmire, G. P. Body size and encephalization in *Homo erectus*. *Anthropos (Brno)* **23**, 139–149 (1986).
- Gould, S. C. Body size of Asian *Homo erectus*: estimation based on prediction models utilizing measures of cranial bone thickness (abstract). *Am. J. Phys. Anthropol.* **16** (suppl.) 93 (1993).
- Hartwig-Scherer, S. body weight prediction in fossil *Homo*. *Cour. Forsch.-Inst. Senckenberg* **171**, 267–279 (1994).
- Ruff, C. B. & Walker, A. in *The Nariokotome Homo Erectus Skeleton* (eds Walker, A. & Leakey, R.) 234–265 (Harvard Univ. Press, Cambridge, 1993).
- Aiello, L. C. & Wood, B. A. Cranial variables as predictors of hominine body mass. *Am. J. Phys. Anthropol.* **95**, 409–426 (1994).
- Kappelman, J. The evolution of body mass and relative brain size in fossil hominids. *J. Hum. Evol.* **30**, 243–276 (1996).
- Ruff, C. B., Scott, W. W. & Liu, A. Y.-C. Articular and diaphyseal remodeling of the proximal femur with changes in body mass in adults. *Am. J. Phys. Anthropol.* **86**, 397–413 (1991).
- Trinkaus, E., Churchill, S. E. & Ruff, C. B. Postcranial robusticity in *Homo*, II: humeral bilateral asymmetry and bone plasticity. *Am. J. Phys. Anthropol.* **93**, 1–34 (1994).
- Grine, F. E., Jungers, W. L., Tobias, P. V. & Pearson, O. M. Fossil *Homo* femur from Berg Aukas, northern Namibia. *Am. J. Phys. Anthropol.* **97**, 151–185 (1995).
- Ruff, C. B. Morphological adaptation to climate in modern and fossil hominids. *Yb. Phys. Anthropol.* **37**, 65–107 (1994).
- Holliday, T. W. *Body Size and Proportions in the Late Pleistocene Western Old World and the Origins of Modern Humans*. (Thesis, Univ. New Mexico, Albuquerque, 1995).
- Wood, B. Origin and evolution of the genus *Homo*. *Nature* **355**, 783–790 (1992).
- Walker, A. in *The Nariokotome Homo Erectus Skeleton* (eds Walker, A. & Leakey, R.) 411–430 (Harvard Univ. Press, Cambridge, 1993).
- Rightmire, G. P. Patterns in the evolution of *Homo erectus*. *Paleobiology* **7**, 241–246 (1981).
- Leigh, S. R. Cranial capacity evolution in *Homo erectus* and early *Homo sapiens*. *Am. J. Phys. Anthropol.* **87**, 1–13 (1992).

- Henneberg, M. Decrease of human skull size in the Holocene. *Hum. Biol.* **60**, 395–405 (1988).
- Frayer, D. W. in *The Origins of Modern Humans: A World Survey of the Fossil Evidence* (eds Smith, F. H. & Spencer, F.) 211–250 (Liss, New York, 1984).
- Tobias, P. V. The negative secular trend. *J. Hum. Evol.* **14**, 347–356 (1985).
- Brown, F., Harris, J., Leakey, R. & Walker, A. Early *Homo erectus* skeleton from West Lake Turkana, Kenya. *Nature* **316**, 788–792 (1985).
- Martin, R. D. *Primate Origins and Evolution* (Princeton Univ. Press, Princeton, 1990).
- Stephan, H., Bauchot, R. & Andy, O. J. in *The Primate Brain* (eds Noback, C. R. & Montague, W.) 289–297 (Appleton-Century-Crofts, New York, 1970).
- Martin, R. D. Relative brain size and basal metabolic rate in terrestrial vertebrates. *Nature* **293**, 57–60 (1981).
- Beals, K. L., Smith, C. L. & Dodd, S. M. Brain size, cranial morphology, climate, and time machines. *Curr. Anthropol.* **25**, 301–330 (1984).
- Hooton, E. A. *The Indians of Pecos Pueblo. A Study of Their Skeletal Remains. Papers of the Phillips Acad. SW Exped.*, No. 4 (Yale Univ. Press, New Haven, 1930).

Acknowledgements. We thank the many institutions and individuals who made available specimens for this study and T. Berger for help in finding the Pecos cranial capacity data. Supported in part by the National Science Foundation and the LSB Leakey Foundation.

Correspondence should be addressed to C.B.R. (e-mail: cruff@welchlink.welch.jhu.edu).

Supplementary Information is on www.nature.com. Paper copies are available from Mary Sheehan at the London editorial office of *Nature*.

Deficits in auditory temporal and spectral resolution in language-impaired children

Beverly A. Wright[†], Linda J. Lombardino[‡], Wayne M. King[‡], Cynthia S. Puranik[‡], Christiana M. Leonard[§] & Michael M. Merzenich^{*}

^{*} *Keck Center for Integrative Neuroscience, Box 0732, University of California, San Francisco, California 94143-0732, USA*

[‡] *Department of Communication Processes and Disorders, § Department of Neuroscience and University of Florida Brain Institute, University of Florida, Gainesville, Florida 32611, USA*

Between 3 and 6 per cent of children who are otherwise unimpaired have extreme difficulties producing and understanding spoken language¹. This disorder is typically labelled specific language impairment. Children diagnosed with specific language impairment often have accompanying reading difficulties (dyslexia)², but not all children with reading difficulties have specific language impairment³. Some researchers claim that language impairment arises from failures specific to language or cognitive processing^{4–6}. Others hold that language impairment results from a more elemental problem that makes affected children unable to hear the acoustic distinctions among successive brief sounds in speech^{7–11}. Here we report the results of psychophysical tests employing simple tones and noises showing that children with specific language impairment have severe auditory perceptual deficits for brief but not long tones in particular sound contexts. Our data support the view that language difficulties result from problems in auditory perception, and provide further information about the nature of these perceptual problems that should contribute to improving the diagnosis and treatment of language impairment and related disorders.

We measured the detection threshold for a brief tone presented before, during or after two different masking noises in eight children diagnosed with specific language impairment, and in eight control children with normal language skills who matched the others in age and non-verbal intelligence (Table 1). Before beginning the tests with the brief tones, we introduced the children to the listening task by measuring their detection thresholds for a long tone presented in the temporal centre of a 'bandpass' noise that included frequencies at and near the tone frequency. The points above the schematic

[†] Present address: Audiology and Hearing Sciences Program, Northwestern University, 2299 North Campus Drive, Evanston, Illinois 60208-3550, USA.

illustration of the stimuli in Fig. 1a show that the same mean tone level was required by specifically language-impaired (filled squares) and control (open squares) children to detect the long tone in this masking condition ($F(1, 14) < 1$, $P > 0.05$).

Figure 1b shows the results of our subsequent measurements in the same children of the detection threshold for a brief tone presented with the bandpass noise at each of four temporal positions shown along the abscissa. The performance pattern of control children (open squares) was just as expected from previous work on normal auditory masking^{12,13}: the tone was easier to detect when it was presented just before or just after, as opposed to during, the noise, and was easiest to detect when it preceded rather than followed the noise. In comparison to controls, children with specific language impairment (filled squares) needed a higher tone level for detection in every condition. Most remarkably, impaired children had as much or more difficulty detecting the tone when it was presented before the noise (the backward-masking condition) as when it was presented during or after the noise. There was no overlap in performance between the two groups in the backward-masking condition.

We also measured detection thresholds in the same children for a brief tone presented at each of the four temporal positions in a spectrally 'notched' noise that excluded frequencies at and near the tone frequency. Figure 2 shows the mean tone thresholds for each group for both the bandpass (squares, replotted from Fig. 1b) and notched (triangles) noises. The conditions are shown at the bottom of Fig. 2. For both impaired and control children, the tone was typically easier to detect with the notched rather than with the bandpass noise. This is expected for normal hearing^{14,15}. However, in contrast to most adults^{16,17}, neither group of children showed a clear threshold difference between the two masker types when the tone and noise started simultaneously.

Two aspects of the performance of impaired children with the notched noise are particularly important. First, language-impaired children were better at hearing the tone presented before (the backward-masking condition) the notched than the bandpass noise. Their severe perceptual deficit for tones presented in this temporal position was worst when the tone and following noise had similar frequencies. Follow-up tests on four additional impaired children showed that the detection threshold in the backward-masking condition reached control values when the spectral notch in the masker was made sufficiently wide. Thus, impaired children had perceptual difficulties in certain temporal and spectral sound contexts, but did not display a general deficit in the perception of rapidly presented sounds. Second, the mean threshold difference

Table 1 Mean of standard scores in the TONI-2 and CELF-R tests

	Language-impaired	Control
Age	8.1 (6.3)	8.0 (7.1)
Sex	6 M, 2 F	3 M, 5 F
TONI-2	101 (5.3)	105.1 (6.5)
CELF-R		
Receptive language	83.4 (6.9)*	112.5 (7.8)
Linguistic concepts	7.4 (1.5)*	10.6 (2.3)
Sentence structure	6.9 (1.6)	12.4 (2.0)
Oral directions	7.5 (2.1)	13.0 (2.1)
Expressive language	70.7 (4.8)*	98.1 (5.2)
Word structure	5.5 (1.6)	8.8 (1.6)
Formulated sentences	4.9 (1.3)	8.6 (1.1)
Recalling sentences	6.3 (2.1)	12.0 (1.3)
Total language	75.3 (4.5)*	105.3 (5.4)

Results in the Test of Nonverbal Intelligence (TONI-2; ref. 25) and in the Clinical Evaluation of Language Fundamentals Revised (CELF-R; ref. 26) are shown for the 8 specifically language-impaired and 8 control children. The mean age (in years and months) and sex distribution of both groups is also shown. The standard deviation of each mean value is given in parentheses.

* Values are based on the mean of 7 of the 8 impaired children as test results were missing.

between the notched and bandpass noises was smaller for impaired than for control children in both the simultaneous-delay (10.5 dB compared to 18.6 dB) and forward (15.7 dB compared to 20.5 dB) masking conditions. This indicates that impaired children were less able than controls to take advantage of a frequency separation between the tone and noise to aid detection of the tone.

A $2 \times 2 \times 4$ analysis of variance performed on the data in Fig. 2 revealed significant main effects for subject group ($F(1, 112) = 102.70$, $P < 0.0001$), noise type ($F(1, 112) = 43.94$, $P < 0.0001$), and tone position ($F(3, 112) = 41.93$, $P < 0.0001$). There were also significant interactions between subject group and tone position ($F(3, 112) = 16.72$, $P < 0.0001$) and noise type and tone position ($F(3, 112) = 3.53$, $P = 0.017$), but not between the subject group and the noise type ($F(1, 112) = 0.08$, $P > 0.05$), nor among the three factors ($F(3, 112) = 1.07$, $P > 0.05$). A Scheffé post hoc analysis indicated that the thresholds of the two subject groups differed significantly only in the two backward-masking conditions ($P < 0.0001$ for the bandpass masker and $P = 0.0002$ for the notched-noise masker).

In summary, these results suggest that children with specific language impairment are severely impaired in their ability to (1) separate a brief sound from a rapidly following sound of similar frequency, and (2) enhance the detection of a brief tone by exploiting a frequency difference between the tone and a longer co-occurring or preceding masking sound. These auditory perceptual deficits could

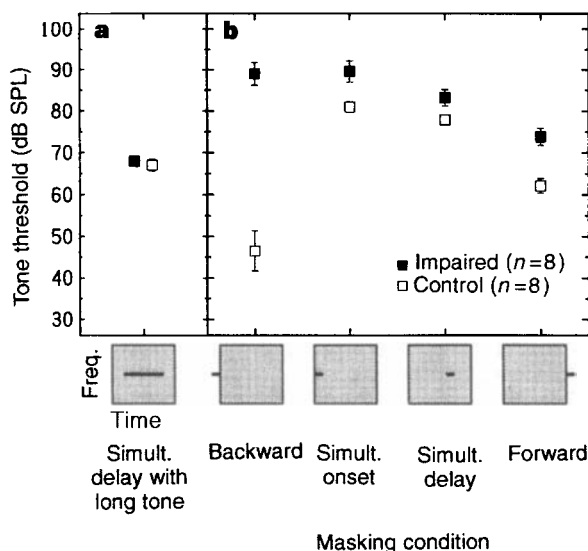


Figure 1 Average tone level required by eight language-impaired (filled squares) and eight control (open squares) children to just detect a long tone temporally centred in a bandpass noise (a), or a brief tone presented before, during or after that noise (b). Error bars indicate plus and minus one standard error of the mean across subjects. The stimuli are represented schematically along the abscissa. SPL, sound pressure level.

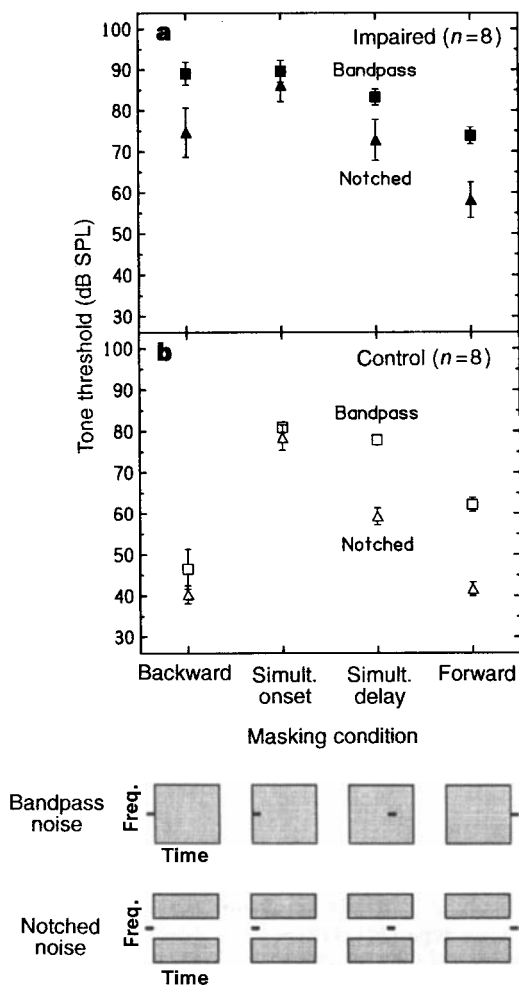


Figure 2 Average tone level required by eight language-impaired (**a**) and eight control (**b**) children to just detect a brief tone presented before, during or after a bandpass (squares, replotted from Fig. 1b) or notched (triangles) noise. Error bars indicate plus and minus one standard error of the mean across subjects. The stimuli are represented schematically along the bottom.

clearly degrade the perception of the brief acoustic elements of speech. Many individuals with language impairment and other disorders related to spoken language might benefit from diagnoses incorporating the auditory tests used here, and from auditory training that focuses on their most severely impaired abilities^{18,19}. The present auditory tests might also aid in the diagnosis and treatment of persons with reading difficulties. We have preliminary data on twelve such individuals. Five had excessive amounts of auditory backward masking, but none had as much masking as the children with specific language impairment. Our results are in accord with the conclusion of a recent review that some, but not all, children with reading problems have difficulties accurately perceiving rapidly presented stimuli²⁰. Our data are also consistent with a previous report showing that children with reading difficulties are particularly poor at discriminating words that differ only in their first sound²¹. Finally, the temporal and spectral specificity of the auditory perceptual deficits reported here may serve to guide the search for the underlying neural bases of language disorders²². □

Methods

Stimuli. All stimuli were generated digitally. Tone: 1,000 Hz, 20 or 200 ms onset-to-offset. Noises: 600–1,400 Hz (bandpass noise) or 400–800 Hz and 1,200–1,600 Hz (notched noise), 300 ms onset-to-offset, 40 dB SPL spectrum level. Gating envelope: 10-ms cosine squared for all stimuli. Masking

conditions: The 20-ms tone was turned on at four different times defined relative to the onset of the 300-ms noise: –20 ms (backward masking), 0 ms (simultaneous-onset masking), 200 ms (simultaneous-delay masking), or 300 ms (forward masking). The 200-ms tone was turned on 50 ms after noise onset.

Procedure. We used a standard, adaptive²³, two-interval, forced-choice procedure to estimate the tone level required for 94% correct detections. The observation intervals were separated by 800 ms. Visual displays on a computer screen marked the observation intervals and gave feedback. Each reported brief-tone threshold was based on the mean of two or three 30-trial measurements per child. Three measurements were always collected, but the most deviant estimate was omitted if the standard deviation of the three was greater than 15 dB. The average within-subject standard error was 3.7 dB for impaired children and 2.5 dB for control children. Because the long-tone condition was used to acquaint the children with the task, we report only the last threshold estimate of the one to three obtained from each child in that condition; in total, the eight impaired children completed 14 threshold estimates and the eight control children 13 estimates in the long-tone condition. We collected the data with the long tone first, and then presented the four brief-tone conditions in pairs (bandpass then notched noise) in a digram balanced latin square²⁴. We tested each child individually in a sound-attenuated room. Stimuli were delivered to the right ear over Sennheiser HD450 headphones. All children were paid for their participation.

Received 18 February; accepted 10 March 1997.

1. Proc. Natl Conf. on Learning Disabilities Bethesda, Maryland, 12 and 13 January 1987 (York, Parkton, MD, 1988).
2. Catts, H. W. The relationship between speech–language impairments and reading disabilities. *J. Speech Hear. Res.* **36**, 948–958 (1993).
3. Aaron, P. G., Kuchta, S. & Grapenthin, C. T. Is there a thing called dyslexia? *Annals Dyslexia* **36**, 33–49 (1988).
4. Aram, D. & Nation, J. Patterns of preschool language disorders. *J. Speech Hear. Res.* **18**, 229–241 (1975).
5. Vellutino, F. R., Steger, B. M., Moyer, S. C., Harding, C. J. & Niles, J. A. Has the perceptual deficit hypothesis led us astray? *J. Learn. Dis.* **10**, 375–385 (1977).
6. Studdert-Kennedy, M. & Mody, M. Auditory temporal perception deficits in the reading-impaired: A critical review of the evidence. *Psychon. Bull. Rev.* **2**, 508–514 (1995).
7. Tallal, P. & Piercy, M. Defects of non-verbal auditory perception in children with developmental aphasia. *Nature* **241**, 468–469 (1973).
8. Frumkin, B. & Rapin, I. Perception of vowels and consonant-vowels of varying duration in language impaired children. *Neuropsychologia* **18**, 443–454 (1980).
9. Lubert, N. Auditory perceptual impairments in children with specific language disorders: A review of the literature. *J. Speech Hear. Dis.* **46**, 3–9 (1981).
10. Elliott, L. L., Hammer, M. A. & Scholl, M. E. Fine-grained auditory discrimination in normal children and children with language-learning problems. *J. Speech Hear. Res.* **32**, 112–119 (1989).
11. Kraus, N. et al. Auditory neurophysiologic responses and discrimination deficits in children with learning problems. *Science* **273**, 971–973 (1996).
12. Fastl, H. Temporal masking effects: II. Critical band noise masker. *Acustica* **36**, 317–330 (1976/77).
13. Soderquist, D. R., Carstens, A. A. & Frank, G. J. H. Backward, simultaneous, and forward masking as a function of signal delay and frequency. *J. Aud. Res.* **21**, 227–245 (1981).
14. Patterson, R. D., Nimmo-Smith, I., Weber, D. L. & Milroy, R. The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold. *J. Acoust. Soc. Am.* **72**, 1788–1803 (1982).
15. Patterson, R. D. & Moore, B. C. J. in *Frequency Selectivity in Hearing* (ed. Moore, B. C. J.) 123–176 (Academic, New York, 1986).
16. Moore, B. C. J., Poon, P. W. F., Bacon, S. P. & Glasberg, B. R. The temporal course of masking and the auditory filter shape. *J. Acoust. Soc. Am.* **81**, 1873–1880 (1987).
17. Carlyon, R. P. Changes in the masked thresholds of brief tones produced by prior bursts of noise. *Hear. Res.* **41**, 223–236 (1989).
18. Merzenich, M. M. et al. Temporal processing deficits of language-learning impaired children ameliorated by training. *Science* **271**, 77–81 (1996).
19. Tallal, P. et al. Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science* **271**, 81–84 (1996).
20. Farmer, M. E. & Klein, R. M. The evidence for a temporal processing deficit linked to dyslexia: A review. *Psychon. Bull. Rev.* **2**, 460–493 (1995).
21. Bradley, L. & Bryant, P. E. Difficulties in auditory organisation as a possible cause of reading backwardness. *Nature* **271**, 746–747 (1978).
22. Gauger, L. M., Lombardino, L. J. & Leonard, C. M. Brain morphology in children with specific language impairment. *J. Speech Hear. Res.* (in the press).
23. Green, D. M. Stimulus selection in adaptive psychophysical procedures. *J. Acoust. Soc. Am.* **87**, 2662–2674 (1990).
24. Wagenaar, W. A. Note on the construction of digram-balanced latin squares. *Psych. Bull.* **72**, 384–386 (1969).
25. Semel, E., Wiig, E. & Secord, W. *The clinical evaluation of language fundamentals revised* (Psychological Corporation, San Antonio, Texas, 1987).
26. Brown, L., Sherbenou, R. J. & Johnson, S. K. *Test of nonverbal intelligence* (Pro-Ed, Austin, Texas, 1990).

Acknowledgements. We thank D. M. Green and C. C. Crandell for access to their laboratories, Z. A. Onsan and Q. T. Nguyen for technical assistance, and B. C. J. Moore for help in improving this manuscript. This work was supported by the McDonnell-Pew Program in Cognitive Neuroscience, the Charles A. Dana Foundation, the March of Dimes, and NIDCD.

Correspondence and requests for materials should be addressed to B.A.W. (e-mail: b-wright@nwu.edu).

Truncation of Kir6.2 produces ATP-sensitive K⁺ channels in the absence of the sulphonylurea receptor

Stephen J. Tucker, Fiona M. Gribble, Chao Zhao, Stefan Trapp & Frances M. Ashcroft

University Laboratory of Physiology, Parks Road, Oxford OX1 3PT, UK

ATP-sensitive potassium channels (K-ATP channels) couple cell metabolism to electrical activity and are important in the physiology and pathophysiology of many tissues¹. In pancreatic β -cells, K-ATP channels link changes in blood glucose concentration to insulin secretion². They are also the target for clinically important drugs such as sulphonylureas, which stimulate secretion, and the K⁺ channel opener diazoxide, which inhibits insulin release^{3,4}. Metabolic regulation of K-ATP channels is mediated by changes in intracellular ATP and Mg-ADP levels, which inhibit and activate the channel, respectively². The β -cell K-ATP channel is a complex of two proteins^{5,6}: an inward-rectifier K⁺ channel subunit, Kir6.2, and the sulphonylurea receptor, SUR1. We show here that the primary site at which ATP acts to mediate K-ATP channel inhibition is located on Kir6.2, and that SUR1 is required for sensitivity to sulphonylureas and diazoxide and for activation by Mg-ADP.

There is accumulating evidence that Kir6.2 acts as the pore-forming subunit of the K-ATP channel, whereas SUR1 acts as a regulator of K-ATP channel activity, conferring sensitivity to sulphonylureas, diazoxide and Mg-ADP⁷⁻¹¹. The mechanism by which ATP inhibits the channel remains unknown and there has been considerable debate about whether the inhibitory ATP-binding site is located on Kir6.2 or SUR1. The facts that most Kir channels do not show marked sensitivity to ATP and that SUR1 possesses two nucleotide-binding domains (NBDs), whereas Kir6.2 has no obvious consensus sites for ATP binding, favour the idea that

SUR1 confers ATP sensitivity on Kir6.2; however, several findings argue against this idea. First, the sensitivity of native K-ATP currents to tolbutamide, diazoxide and Mg-ADP declines following patch excision, whereas the ability of ATP to block the channel remains unimpaired². Likewise, mild proteolysis of the inner membrane surface abolishes the sulphonylurea and Mg-ADP sensitivity of native K-ATP channels, but only slightly reduces their ATP sensitivity¹². Both these results suggest that Kir6.2 may functionally dissociate from SUR1 yet still retain sensitivity to ATP. The problem of which K-ATP channel subunit possesses the ATP-inhibitory site has been difficult to address because, unlike most other Kir channels, Kir6.2 does not express functional channels alone, but only when coexpressed with a sulphonylurea receptor (SUR1 or SUR2; refs 5, 6, 8). The reason for this is unclear. Although Kir6.2 is stably expressed in the absence of SUR1 (ref. 13), it may not reach the cell surface. An alternative idea is that Kir6.2 is present in the plasma membrane but in an inactive state. As the amino terminus of voltage-gated K⁺ channels acts as a blocking particle, plugging the pore and inactivating the channel¹⁴, we reasoned that the same might be true for Kir6.2. We therefore tested the effects of deleting the distal parts of the amino or carboxy terminus of Kir6.2.

As previously reported¹³, large whole-cell currents can be recorded in response to metabolic inhibition from *Xenopus* oocytes coinjected with mRNAs encoding wild-type Kir6.2 and SUR1, but not from oocytes injected with wild-type Kir6.2 mRNA alone (Fig. 1a, b). In contrast, significant currents were recorded from oocytes injected with mRNA encoding a truncated form of Kir6.2 in which either the last 26 (Kir6.2 Δ C26) or 36 (Kir6.2 Δ C36) amino acids of the C terminus had been deleted. These currents were further enhanced by metabolic inhibition (Fig. 1a, b). This suggests that Kir6.2 Δ C26 and Kir6.2 Δ C36 are not only capable of independent expression but are also intrinsically sensitive to metabolically induced changes in cytosolic nucleotide levels.

To explore this possibility, we recorded macroscopic currents from giant inside-out patches excised from *Xenopus* oocytes expressing either wild-type Kir6.2, or Kir6.2 with N- or C-terminal deletions of different lengths. A marked increase in current was observed when patches were excised into nucleotide-free solution from oocytes expressing either Kir6.2 Δ C26 or Kir6.2 Δ C36 (Fig. 1c). This reflects the relief of the blocking action of ATP present in the

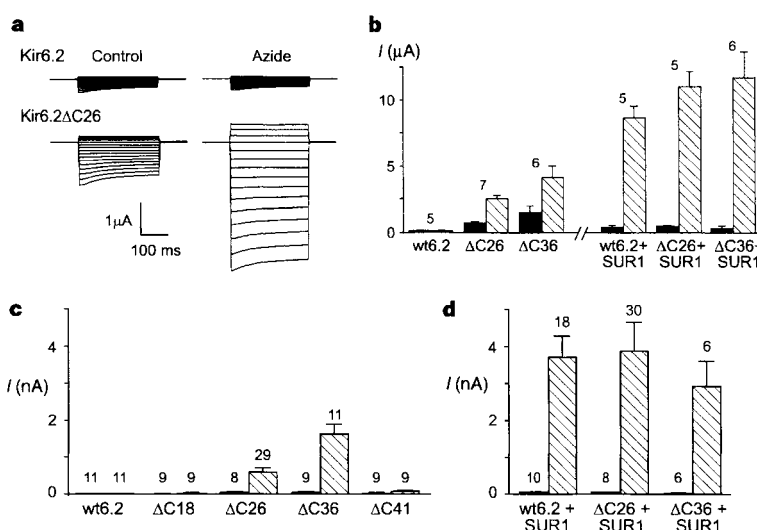


Figure 1 a, Whole-cell currents recorded from an oocyte injected with mRNA encoding wild-type (wt) Kir6.2 or Kir6.2 Δ C26 in response to a series of voltage steps from -120 to +20 mV before and 15 min after exposure to 3 mM azide. **b**, Mean whole-cell currents recorded at -100 mV before (filled bars) and 15 min after (hatched bars) exposure to 3 mM azide from oocytes injected with mRNA encoding wtKir6.2, Kir6.2 Δ C26, Kir6.2 Δ C36, wtKir6.2 + SUR1 (1:1 ratio),

Kir6.2 Δ C26 + SUR1 (1:1 ratio) or Kir6.2 Δ C36 + SUR1 (1:50 ratio). The number of oocytes is given above the bars. **c, d**, Mean macroscopic current amplitudes at -100 mV before (filled bars) and after (hatched bars) patch excision from oocytes injected with the mRNAs indicated. The number of patches is given above the bars.

oocyte cytoplasm. No increase in current on patch excision was observed with oocytes injected with wild-type Kir6.2 or when 14 residues were deleted from the N terminus of Kir6.2 ($n = 9$), and only a very small increase (<100 pA) was detected when either the last 18 residues (Kir6.2 Δ C18) or 41 residues (Kir6.2 Δ C41) of Kir6.2 were deleted from the C terminus (Fig. 1c). These results indicate that the last 18, or possibly 26, amino acids of Kir6.2 can prevent its independent functional expression. Coexpression of SUR1 enabled wild-type Kir6.2 to be expressed and also enhanced both Kir6.2 Δ C26 and Kir6.2 Δ C36 currents (Fig. 1d).

It is unclear why deletion of the C terminus of Kir6.2 should allow functional expression of the protein. One possibility is that the C terminus inhibits channel activity by blocking the pore. If so, then coupling of SUR1 to Kir6.2 might prevent the action of this blocking particle and so permit channel function. Alternatively, the C terminus might interact with other proteins that prevent either functional activity or surface expression of Kir6.2, effects that would be reversed by SUR1. The ability to express the truncated form of

Kir6.2 independently of SUR1 allowed us to investigate which properties of the K-ATP channel are intrinsic to Kir6.2 and which are endowed by SUR1.

Figure 2a, b shows single-channel currents and mean single-channel current–voltage relations recorded from inside-out patches excised from oocytes expressing Kir6.2 Δ C26 or Kir6.2 Δ C26 plus SUR1. The single-channel conductance, measured between -20 and -80 mV, was 69 ± 2 pS ($n = 4$) for Kir6.2 Δ C26 and 73 ± 2 pS ($n = 5$) for Kir6.2 Δ C26 plus SUR1. Thus, as SUR1 does not modify the single-channel conductance, it probably does not contribute to the main part of the permeation pathway. The single-channel conductance of Kir6.2 Δ C36 was 75 ± 2 pS ($n = 5$). None of these values are significantly different from those of wild-type Kir6.2/SUR1 currents (70 – 76 pS; refs 5, 6) or native K-ATP currents (50 – 80 pS; ref. 2), indicating that deletion of the C terminus does not alter the single-channel conductance.

Application of ATP to the inner membrane surface markedly inhibited both Kir6.2 Δ C26 and Kir6.2 Δ C36 currents (Fig. 2c, d), with half-maximal block (K_i) at 106 ± 4 μ M ($n = 7$) and 128 ± 5 μ M ($n = 6$), respectively. This suggests the Kir6.2 subunit possesses an intrinsic ATP-inhibitory site. The Hill coefficient (h) was 1.2 ± 0.1 ($n = 7$) for Kir6.2 Δ C26 and 1.0 ± 0.1 ($n = 6$) for Kir6.2 Δ C36. Its value is close to unity, which indicates that only a single ATP molecule needs to bind to close the channel. In β -cells, ATP inhibition does not require hydrolysis of the nucleotide². This was also the case for Kir6.2 Δ C36 currents as ATP-inhibition was unaffected by Mg^{2+} removal: 100 μ M ATP blocked the slope conductance by $44 \pm 2\%$ ($n = 9$) in the presence of 1.4 mM Mg^{2+} and by $42 \pm 2\%$ ($n = 6$) in Mg^{2+} -free solution. Coexpression of SUR1 with Kir6.2 Δ C26 increased the potency of ATP inhibition (Fig. 2c, d) to that found for wild-type K-ATP channels: $K_i = 13.4 \pm 0.3$ μ M ($n = 8$) for Kir6.2 Δ C26/SUR1 and 10 – 28 μ M for wild-type Kir6.2/SUR1 currents^{5,13}. Likewise, the K_i for ATP inhibition of Kir6.2 Δ C36/SUR1 currents (24.6 ± 2.4 μ M; $n = 5$) was lower than that of Kir6.2 Δ C36 currents (128 μ M).

To exclude the possibility that Kir6.2 Δ C36 or Kir6.2 Δ C26 couple to an endogenous protein in the *Xenopus* oocyte that makes it ATP-sensitive, we also examined whether truncated Kir6.2 expresses ATP-sensitive currents in mammalian cells. Whole-cell currents recorded from HEK293 cells transfected with Kir6.2 Δ C36 increased

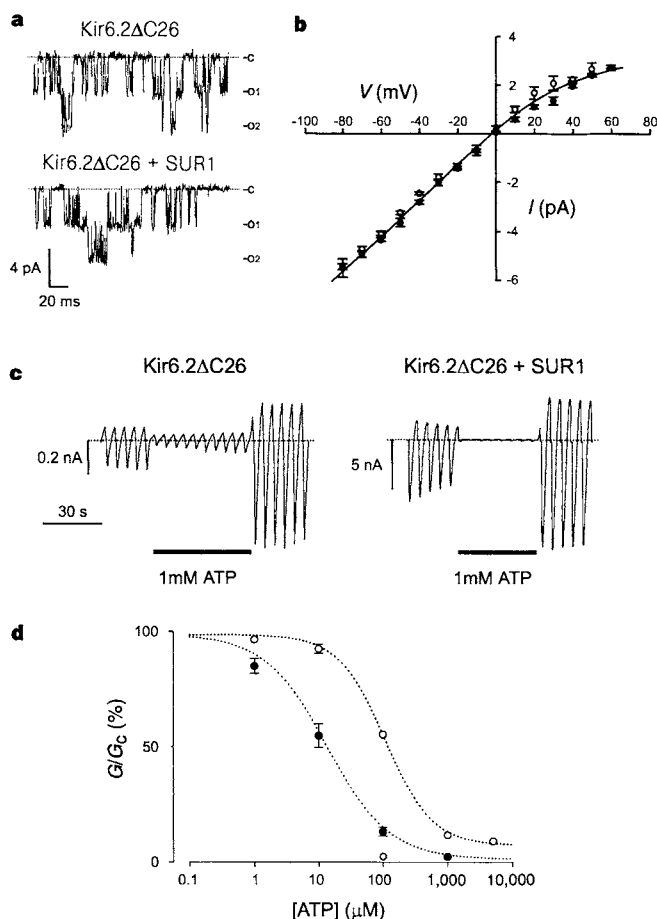


Figure 2 a, Single-channel currents recorded at -60 mV from an inside-out patch excised from an oocyte injected with mRNA encoding Kir6.2 Δ C26, or Kir6.2 Δ C26 + SUR1. b, Mean single-channel current–voltage relations recorded for Kir6.2 Δ C26 (circles) or Kir6.2 Δ C26 + SUR1 (filled circles). c, Macroscopic currents recorded from 2 different inside-out patches in response to a series of voltage ramps from -110 to $+100$ mV. Oocytes were injected with mRNAs encoding Kir6.2 Δ C26 or Kir6.2 Δ C26 + SUR1. 1 mM ATP was added to the internal solution as indicated by the bar. d, Mean ATP dose–response relationships for Kir6.2 Δ C26 currents (open circles; $n = 7$) and Kir6.2 Δ C26/SUR1 currents (filled circles; $n = 6$). Test solutions were alternated with control solutions and the slope conductance (G) is expressed as a percentage of the mean (G_c) of that in the control solution before and after exposure to ATP. Conductance was measured between -20 and -100 mV and is the mean of 5 voltage ramps. The lines are the best fits of the data to the Hill equation using the mean values for K_i and h given in the text.

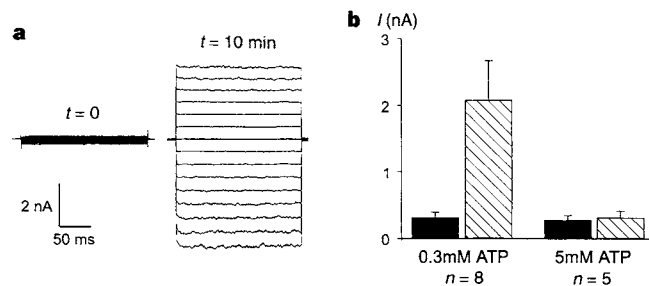


Figure 3 a, Whole-cell currents recorded from an HEK293 cell transfected with cDNA encoding Kir6.2 Δ C36 in response to a series of voltage steps from -110 to $+30$ mV, immediately after forming the whole-cell configuration and ~ 10 min later. The pipette contained 0.3 mM ATP. b, Mean whole-cell current amplitudes recorded from HEK293 cells transfected with cDNA encoding Kir6.2 Δ C36 immediately after forming the whole-cell configuration (black bars) and ~ 10 min later (hatched bars). Cells were dialysed with the ATP concentrations indicated.

with time following establishment of the whole-cell configuration when dialysed intracellularly with 0.3 mM ATP (Fig. 3). This increase in current results from the washout of endogenous ATP from the cell as it was not evident when cells were dialysed with 5 mM ATP. Similar results were found for Kir6.2ΔC26 currents (not shown) and wild-type Kir6.2/SUR1 currents⁶. Other K⁺ channels that require an additional subunit for expression in oocytes (for example, Kir3.1 and Isk) are not expressed by themselves in mammalian cells, but in *Xenopus* oocytes they give small currents that are limited in size by the amount of endogenous subunit present^{15,16}. This is not the case for Kir6.2ΔC26 and Kir6.2ΔC36, strongly suggesting that they are capable of independent expression and so are intrinsically ATP sensitive.

As the free ATP molecule carries four negative charges, it seems likely that a positively charged residue may form part of the ATP-binding site. We found that neutralization of the lysine residue at position 185 to glutamine (Kir6.2ΔC26-K185Q) substantially decreased the channel sensitivity to ATP (Fig. 4): K_i was 4.2 ± 0.2 mM ($n = 6$) for Kir6.2ΔC26-K185Q currents compared with 106 ± 4 μM ($n = 7$) for Kir6.2ΔC26 currents. The Hill coefficient was unaffected, being 1.2 ± 0.1 for both Kir6.2ΔC26-K185Q ($n = 6$) and Kir6.2ΔC26 ($n = 7$) currents. These data provide further support for the idea that ATP interacts directly with Kir6.2 to mediate channel inhibition. They also suggest that Lys 185 is involved in this inhibition, either by forming part of the ATP-binding site itself or by constituting part of the transduction mechanism by which binding of ATP to a site elsewhere effects channel closure.

Figure 2d shows that although the primary site at which ATP inhibits the K-ATP channel resides on Kir6.2, the presence of SUR1 can enhance the blocking action of the nucleotide (from ~100 to ~10 μM). Furthermore, when SUR1 was coexpressed with Kir6.2ΔC26-K185Q, the ATP sensitivity was also slightly greater than that found in the absence of the sulphonylurea receptor ($K_i = 2.3 \pm 0.1$ mM; $n = 6$, $P < 0.001$). This enhancement need not imply that ATP interacts with a second inhibitory site on SUR1: for example, the sulphonylurea receptor may simply facilitate access of ATP to its inhibitory site on Kir6.2. One reason it has been postulated that the inhibitory binding site for ATP lies on SUR1 is that the K_i for ATP inhibition of Kir6.2/SUR1 currents is ~10 μM, whereas that for Kir6.2/SUR2 currents is ~100 μM (refs 5, 8). Our results suggest this difference may be because SUR1, but not SUR2, enhances the ATP sensitivity of Kir6.2.

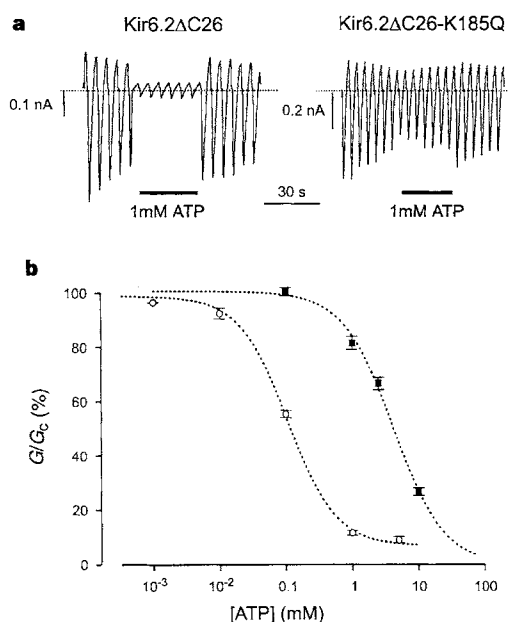


Figure 4 a, Macroscopic currents recorded from 2 different inside-out patches in response to a series of voltage ramps from -110 to +100 mV. Oocytes were injected with mRNAs encoding Kir6.2ΔC26 or Kir6.2ΔC26-K185Q. 1 mM ATP was added to the internal solution as indicated by the bar. **b**, Mean ATP dose-response relationships for Kir6.2ΔC26 currents (open circles; $n = 7$) and Kir6.2ΔC26-K185Q currents (filled squares; $n = 6$). Test solutions were alternated with control solutions and the slope conductance (G) is expressed as a percentage of the mean (G_c) of that obtained in control solution before and after exposure to ATP. Conductance was measured between -20 and -100 mV and is the mean of 5 voltage ramps. The lines are the best fits of the data to the Hill equation using the mean values for K_i and the Hill coefficient given in the text.

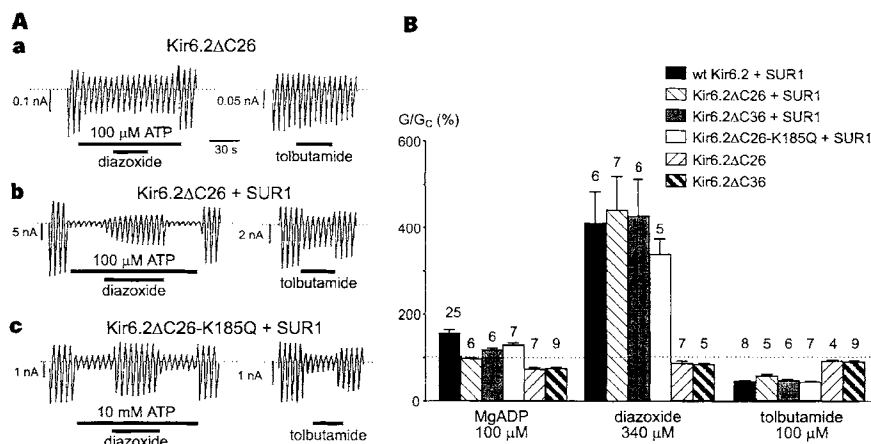


Figure 5 A, Macroscopic currents recorded from inside-out patches in response to a series of voltage ramps from -110 to +100 mV. Oocytes were injected with mRNAs encoding Kir6.2ΔC26 (**a**), Kir6.2ΔC26 + SUR1 (1:1 ratio) (**b**), or Kir6.2ΔC26-K185Q + SUR1 (1:50 ratio) (**c**). 100 μM Mg-ATP, 340 μM diazoxide or 100 μM tolbutamide was added to the internal solution, as indicated by the bars. **B**, Mean macroscopic slope conductance recorded in the presence of 100 μM

Mg-ADP or 100 μM tolbutamide, expressed as a percentage of the slope conductance in control solution (no additions). Diazoxide (340 μM) was added in the presence of 100 μM Mg-1-ATP or 10 mM Mg-ATP (Kir6.2ΔC26-K185Q + SUR1, only) and is expressed as a percentage of the current amplitude in Mg-ATP solution. The dashed line indicates the current level in the absence of the test compound. The number of patches is given above the bars.

The finding that the ATP-inhibitory site is located on Kir6.2, rather than on the sulphonylurea receptor, is consistent with the facts that SUR2B forms an ATP-sensitive K^+ channel with Kir6.2 but an ATP-insensitive K^+ channel with Kir6.1 (ref. 17), and that other Kir subunits such as Kir1.1b (ROMK2) and Kir2.3, which are expressed independently, are blocked by high concentrations of ATP (~ 5 mM; refs 18, 19). Furthermore, the reported mutations within the NBDs of SUR1 do not reduce ATP sensitivity^{10,11}. It is perhaps not surprising that the NBDs of SUR1 do not act as the primary site of ATP block because in other ABC transporters the primary role of these domains is to hydrolyse Mg-ATP²⁰, whereas the β -cell K-ATP channel is blocked by the free base ATP⁴⁻ (refs 1, 2).

As found for the native and wild-type K-ATP currents^{2,11}, both Kir6.2AC26 (Fig. 2) and Kir6.2AC36 currents ran down with time in excised patches, and were refreshed following exposure to Mg-ATP. This indicates that both these properties are intrinsic to Kir6.2 rather than to SUR1.

Both native β -cell K-ATP and wild-type Kir6.2/SUR1 currents are blocked by tolbutamide and potentiated by Mg-ADP and diazoxide^{1,5,6}; the stimulatory action of the latter is dependent upon the presence of intracellular hydrolysable nucleotides^{1,2}. Figure 5A, a, shows that both tolbutamide and diazoxide (in the presence of 100 μ M ATP) were without effect on Kir6.2AC26 currents at concentrations that markedly alter wild-type Kir6.2/SUR1 currents¹³. Furthermore, Mg-ADP blocked Kir6.2AC26 currents (Fig. 5B). This result is not unexpected, because the NBDs of SUR1 confer Mg-ADP activation on the K-ATP channel and in the absence of this activation an inhibitory effect of the nucleotide is unmasked^{10,11}. It also suggests that, like ATP, the inhibitory effect of ADP is intrinsic to Kir6.2. Metabolic inhibition caused greater activation of whole-cell Kir6.2AC26/SUR1 (or Kir6.2AC36/SUR1) currents than whole-cell Kir6.2AC26 (or Kir6.2AC36) currents (Fig. 1b). This may reflect the potentiatory action of Mg-ADP conferred by SUR1 and supports the view that the stimulatory effect of Mg-ADP is important in coupling metabolism to K-ATP channel activity^{1,2,10,11}.

Coexpression of Kir6.2AC26 with SUR1 restored the efficacy of diazoxide to that found for wild-type currents and almost fully restored the effects of tolbutamide and Mg-ADP (Fig. 5A, b and B). These data confirm that SUR1 makes the K-ATP channel sensitive to sulphonylureas, diazoxide and the potentiatory action of Mg-ADP [7–11]. Similar results were found for Kir6.2AC36, providing that an excess of SUR1 mRNA was injected (Fig. 5B); when equivalent amounts of mRNAs were injected, no coupling was observed. Likewise, coexpression of Kir6.2AC26-K185Q with SUR1 resulted in currents that were inhibited by tolbutamide and activated by Mg-ADP and diazoxide (Fig. 5A, c and B). This indicates that although K185 is involved in channel closure induced by ATP, it is not required for inhibition by sulphonylureas.

In conclusion, we have shown that the last 18 (or 26) amino acids of Kir6.2 prevent its independent functional expression. This effect is reversed by SUR1. Our results provide evidence that Kir6.2 acts as the pore of the K-ATP channel complex and that SUR1 endows Kir6.2 with sensitivity to sulphonylureas, diazoxide and Mg-ADP. Importantly, they also indicate that the primary site at which ATP interacts to mediate channel inhibition is located on Kir6.2 and not on SUR1, as previously suggested. However, although the ATP-inhibitory site lies on Kir6.2, SUR1 enhances the sensitivity of Kir6.2 to ATP (from $K_i \approx 100$ μ M to $K_i \approx 10$ μ M). Our results are likely to be applicable to K-ATP channels in brain, cardiac and skeletal muscle, where Kir6.2 is also thought to act as the pore-forming subunit^{5,6,8,21}. Finally, that Kir6.2AC26 can be expressed independently makes it a useful tool for determining whether drugs that act on K-ATP channels do so by interacting with Kir6.2 or with SUR1, and for examining the functional interaction between Kir6.2 and SUR1. □

Methods

Molecular biology. C-terminal deletions of mouse Kir6.2 (Genbank D50581) were made by introduction of a stop codon at the appropriate residues by site-directed mutagenesis. The N-terminal deletion was similarly constructed by removal of the original start codon and replacement of the leucine residue at position 14 with an initiator methionine. Synthesis of mRNA encoding wild-type and mutant mouse Kir6.2 and wild-type rat SUR1 (Genbank, L40624) was performed as described¹³.

Electrophysiology. *Xenopus* oocytes were defolliculated and injected with ~ 2 ng mRNA encoding either full-length or truncated Kir6.2. In coexpression experiments, ~ 2 ng wild-type (wt) Kir6.2, ~ 2 ng Kir6.2AC26 or ~ 0.04 ng Kir6.2AC36 mRNA was injected with 2 ng SUR1 mRNA (giving a 1:1 or 1:50 ratio). The final injection volume was ~ 50 nl per oocyte. Isolated oocytes were maintained in tissue culture¹³ and studied 1–4 days after injection. Whole-cell currents were measured at 18–24 °C using a standard 2-electrode voltage clamp¹³ in (mM): 90 KCl, 1 MgCl₂, 1.8 CaCl₂, 5 HEPES (to pH 7.4 with KOH). The holding potential was -10 mV. Currents were filtered at 1 kHz, digitized at 4 kHz and measured 280–295 ms after the start of the voltage pulse. Macroscopic (or single-channel) currents were recorded from giant (or normal) excised inside-out patches at a holding potential of 0 mV and at 20–24 °C (refs 11, 22). The pipette solution contained (mM): 140 KCl, 1.2 MgCl₂, 2.6 CaCl₂, 10 HEPES (pH 7.4 with KOH) and the internal (bath) solution contained (mM): 110 KCl, 1.44 MgCl₂, 30 KOH, 10 EGTA, 10 HEPES (pH 7.2 with KOH), plus nucleotides as indicated. The Mg²⁺-free solution contained (mM): 110 KCl, 30 KOH, 2.6 CaCl₂, 10 EDTA, 10 HEPES (pH 7.2 with KOH). Macroscopic currents were recorded in response to 4-s voltage ramps from -110 to $+100$ mV with an EPC7 amplifier (List Elektronik, Darmstadt, Germany), filtered at 0.2 kHz and sampled at 0.5 kHz. The slope conductance was measured by fitting a straight line to the data between -20 and -100 mV; the average of 5 consecutive ramps was calculated in each solution. Single-channel currents were filtered at 1 kHz and sampled at 3 kHz.

HEK293 cells were transiently transfected with the pcDNA3 vector (Invitrogen) containing the coding sequence of wild-type or mutant Kir6.2 using lipofectamine (GibcoBRL)⁶. Whole-cell currents were studied 48–72 h after transfection. The pipette solution contained (mM): 107 KCl, 1.2 MgCl₂, 1 CaCl₂, 10 EGTA, 5 HEPES (pH 7.2 with KOH; total $K^+ \sim 140$ mM) and either 0.3 or 5 mM ATP. The bath solution contained (mM): 40 KCl, 100 NaCl, 2.6 CaCl₂, 1.2 MgCl₂, 5 HEPES (pH 7.4). The holding potential was -30 mV.

Data analysis. All data is given as mean $\pm$ s.e.m. The symbols in the figures indicate the mean and the vertical bars one s.e.m. (where this is larger than the symbol). ATP dose–response relationships were fitted to the Hill equation: $G/G_c = O + (100 - O)/(1 + ([ATP]/K_i)^h)$, where [ATP] is the ATP concentration, K_i is the ATP concentration at which inhibition is half maximal, h is the slope factor (Hill coefficient) and O is the background current (not blocked by ATP). Statistical significance was tested using Student's *t*-test.

Received 14 January; accepted 21 March 1997.

1. Ashcroft, F. M. & Ashcroft, S. J. H. Properties and functions of ATP-sensitive K-channels. *Cell. Signal* **2**, 197–214 (1990).
2. Ashcroft, F. M. & Rorsman, P. Electrophysiology of the pancreatic β -cell. *Prog. Biophys. Mol. Biol.* **54**, 87–143 (1989).
3. Ashcroft, F. M. & Ashcroft, S. J. H. The sulphonylurea receptor. *Biochim. Biophys. Acta* **1175**, 45–59 (1992).
4. Dunne, M. J., Harding, E., Jaggar, J. H., Ayton, B. J. & Squires, P. E. in *Frontiers of Insulin Secretion and Pancreatic β -Cell Research*, (eds Flatt, P. & Lenzen, S.) 153–59 (Smith Gordon, UK, 1993).
5. Inagaki, N. et al. Reconstitution of IKATP: an inward rectifier subunit plus the sulphonylurea receptor. *Science* **270**, 1166–1169 (1995).
6. Sakura, H., Åmmälä, C., Smith, P. A., Gribble, F. M. & Ashcroft, F. M. Cloning and functional expression of the cDNA encoding a novel ATP-sensitive potassium channel expressed in pancreatic β -cells, brain, heart and skeletal muscle. *FEBS Lett.* **377**, 338–344 (1995).
7. Anguilar-Bryan, L. et al. Cloning of the β -cell high-affinity sulphonylurea receptor: a regulator of insulin secretion. *Science* **268**, 423–425 (1995).
8. Inagaki, N. et al. A family of sulphonylurea receptors determines the properties of ATP-sensitive K^+ channels. *Neuron* **16**, 1011–1017 (1996).
9. Åmmälä, C., Moorhouse, A. & Ashcroft, F. M. The sulphonylurea receptor confers diazoxide sensitivity on the inwardly-rectifying K-channel, Kir6.1. *J. Physiol. (Lond.)* **494**, 709–714 (1996).
10. Nichols, C. G. et al. Adenosine diphosphate as an intracellular regulator of insulin secretion. *Science* **272**, 1785–1787 (1996).
11. Gribble, F. M., Tucker, S. J. & Ashcroft, F. M. The essential role of the Walker A motifs of SUR1 in K-ATP channel activation by MgADP and diazoxide. *EMBO J.* **16**, 1145–1152 (1997).
12. Proks, P. & Ashcroft, F. M. Modification of K-ATP channels in pancreatic β -cells by trypsin. *Pflügers Arch.* **424**, 63–72 (1993).
13. Gribble, F. M., Ashfield, R., Åmmälä, C. & Ashcroft, F. M. Properties of cloned ATP-sensitive K-currents expressed in *Xenopus* oocytes. *J. Physiol. (Lond.)* **498**, 87–98 (1997).
14. Hoshi, T., Zagotta, W. N. & Aldrich, R. W. Biophysical and molecular mechanisms of Shaker potassium channel inactivation. *Science* **250**, 533–538 (1990).

15. Krapivinsky, G. *et al.* The G-protein-gated atrial K⁺ channel I_{KACH} is a heteromultimer of two inwardly rectifying K⁺-channel proteins. *Nature* **374**, 135–141 (1995).
16. Sanguinetti, M. C. *et al.* Coassembly of K_vLQT1 and MinK (IsK) proteins to form cardiac I_{Ks} potassium channel. *Nature* **384**, 80–83 (1996).
17. Yamada, Y. M. *et al.* Sulphonylurea receptor 2B and Kir6.1 form a sulphonylurea-sensitive but ATP insensitive K⁺ channel. *J. Physiol. (Lond.)* **499**, 715–720 (1997).
18. McNicholas, C. M., Yang, Y., Gebeisch, G. & Hebert, S. C. Molecular site for nucleotide binding on an ATP-sensitive renal K⁺ channel (ROMK2). *Am. J. Physiol.* **271**, F275–F285 (1996).
19. Collins, A., German, M. S., Jan, Y. N., Jan, L. Y. & Zhao, B. A strongly inwardly rectifying K⁺ channel that is sensitive to ATP. *J. Neurosci.* **16**, 1–9 (1996).
20. Higgins, C. F. ABC transporters: from microorganisms to man. *Annu. Rev. Cell Biol.* **8**, 67–113 (1992).
21. Karschin, C., Ecke, C., Ahse, F. M. & Karschin, A. Overlapping tissue distribution of K_{ATP} channel-forming Kir6.2 subunit and the sulphonylurea receptor SUR1 in rodent brain. *FEBS Lett.* **401**, 59–64 (1996).
22. Hilgemann, D. W., Nicoll, D. A. & Philipson, K. D. Charge movement during Na⁺ translocation by native and cloned Na⁺/Ca²⁺ exchanger. *Nature* **352**, 715–718 (1991).

Acknowledgements. We thank G. Bell for the gift of rat SUR1, and B. Fakler and P. Ruppersberg for discussion. We also thank the Wellcome Trust, the British Diabetic Association and the MRC for support. F.M.G. is an MRC Clinical Training Fellow, S.T. is supported by the Deutsche Forschungsgemeinschaft, and S.J.T. is a Wellcome Trust Fellow.

Correspondence and requests for materials should be addressed to F.M.A. (e-mail: frances.ashcroft@physiol.ox.ac.uk).

Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection

Tae-Wook Chun*, **Lucy Carruth***, **Diana Finzi***, **Xuefei Shen***, **Joseph A. DiGiuseppe†**, **Harry Taylor***, **Monika Hermankova***, **Karen Chadwick‡**, **Joseph Margolick‡**, **Thomas C. Quinn§**, **Yen-Hong Kuol**, **Ronald Brookmeyer||**, **Martha A. Zeiger†**, **Patricia Barditch-Crovo*** & **Robert F. Siliciano***

Departments of *Medicine, †Pathology, and ‡Surgery, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

Departments of §Molecular Microbiology and Immunology, and ||Biostatistics, Johns Hopkins University School of Hygiene and Public Health, Baltimore, Maryland 21205, USA

§ National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

The capacity of HIV-1 to establish latent infection of CD4⁺ T cells may allow viral persistence despite immune responses and antiretroviral therapy. Measurements of infectious virus^{1,2} and viral RNA^{3,4} in plasma and of infectious virus¹, viral DNA^{5–10} and viral messenger RNA species^{11–14} in infected cells all suggest that HIV-1 replication continues throughout the course of infection. Uncertainty remains over what fraction of CD4⁺ T cells are infected and whether there are latent reservoirs for the virus. We show here that during the asymptomatic phase of infection there is an extremely low total body load of latently infected resting CD4⁺ T cells with replication-competent integrated provirus (<10⁷ cells). The most prevalent form of HIV-1 DNA in resting and activated CD4⁺ T cells is a full-length, linear, unintegrated form that is not replication competent. The infection progresses even though at any given time in the lymphoid tissues integrated HIV-1 DNA is present in only a minute fraction of the susceptible populations, including resting and activated CD4⁺ T cells and macrophages.

The importance of latent reservoirs for HIV-1 is highlighted by studies of potent antiretroviral drugs that block new rounds of infection and produce a dramatic drop in plasma virus in two weeks. The rapid decline shows that virus production results largely from continuous rounds of virus infection of and replication in host cells with rapid turnover of free virus and virus-producing cells^{15–18}. Subsequently, however, plasma virus declines at a slower rate, presumably reflecting the turnover of long-term reservoirs of chronically or latently infected cells³⁰. Latent infection of CD4⁺ T cells may occur through two mechanisms, both of which are operative only in cells that are in a resting G₀ state. Partially or

completely reverse-transcribed HIV-1 DNA may be present in the cytoplasm of infected resting CD4⁺ T cells as a result of a block in reverse transcription or nuclear import and may constitute a labile, inducible reservoir^{6,19,20}. A second, more stable latent reservoir comprises resting CD4⁺ T cells with integrated provirus²¹. Little is known about the size, characteristics and functional significance of latent cellular reservoirs for HIV-1.

The latent reservoirs were studied in 14 asymptomatic, HIV-1-infected donors (CD4 counts 200–700 cells μl⁻¹; plasma RNA 1,400–116,000 copies ml⁻¹). Half of the donors were receiving antiretroviral therapy (nucleoside analogues). Because lymph nodes may represent an important site for viral replication^{9,10}, node biopsies were obtained. These showed follicular hyperplasia, follicular involution, and mixed patterns. Immunohistochemical staining for p24 antigen revealed diffuse follicular localization and rare productively infected cells with intense cytoplasmic staining. To determine the frequency of latently infected CD4⁺ T cells with integrated HIV-1 DNA, we isolated resting CD4⁺ T cells from lymph nodes and blood. Although the fraction of CD4⁺ T cells that were activated (HLA DR⁺) was higher in the lymph nodes than in blood (40 ± 13% versus 24 ± 9%), resting CD4⁺ T cells could be isolated from both sources with >99% purity. Resting cells did not release detectable virus. The frequency of resting CD4⁺ T cells with integrated HIV-1 DNA was measured using a quantitative version of an inverse polymerase chain reaction (PCR) assay²¹ that specifically detects integrated HIV-1 DNA by amplifying unique upstream host genomic sequences (Fig. 1a). Sequence analysis of the inverse PCR products revealed fusion of the 5' long terminal repeat (LTR) to cellular (non-HIV) sequences, with the expected loss of two bases from the HIV-1 genome at the junction²² and ligation of gag sequences to cellular DNA at the PstI site (Fig. 1b). The inverse PCR assay was carried out at a limiting dilution with a second nested amplification of an internal gag sequence. This approach compensates for the tendency of inverse PCR to give preferential amplification of small templates. For a control plasmid construct and for ACH-2 cells, which give inverse PCR products of 0.4 and 1.7 kilobases (kb), respectively, very strong ligase-dependent bands were seen even after dilution to the level of 3–5 copies in 30,000 cell equivalents of DNA (Fig. 1c). The signal extinguished upon dilution to 0.3–0.5 copies. Thus, for integration events giving inverse PCR bands of <2 kb, the assay can detect a single integration in 30,000 cell equivalents of human DNA. Because dilution series

Table 1 Virus load in activated CD4⁺ T cells

Donor*	CD4 ⁺ DR ⁺ T cells with integrated HIV† (per 10 ⁶ cells)	Upper bound for CD4 ⁺ DR ⁺ T cells with integrated HIV‡ (per 10 ⁶ cells)	Total body load of CD4 ⁺ DR ⁺ T cells with integrated HIV§
4	387	1,410	5.2 × 10 ⁷
5	101	410	1.8 × 10 ⁷
6	257	1,280	2.3 × 10 ⁷
8	172	542	2.5 × 10 ⁷
9	598	1,768	6.7 × 10 ⁷
10	345	1,132	5.9 × 10 ⁷
12	80	376	1.2 × 10 ⁷
Mean	224	—	3.1 × 10 ⁷

* Calculations were made for donors whose lymph node samples had clearly measurable integrated HIV-1 DNA in both the unfractionated and purified resting CD4⁺ T cell populations. In the remaining donors, integrated HIV-1 DNA was below the limit of detection in the unfractionated populations, suggesting an even lower total body load of activated CD4⁺ T cells with integrated HIV-1 DNA.

† Determined by subtraction using values of integrated HIV-1 DNA in unfractionated and purified resting CD4⁺ T cells and the frequencies of resting and activated CD4⁺ T cells in the lymph nodes. Calculations are based on the assumption that after depletion of adherent cells, all integrated HIV-1 in lymph node mononuclear cells is either in resting or activated CD4⁺ T cells.

‡ Because these estimates are based on the subtraction of two experimentally derived values, each with its own variance, an upper bound on the frequency of activated CD4⁺ T cells with integrated HIV-1 DNA was also determined as the calculated frequency +2s.e. Standard errors were determined using the rule of total probability from the variances for the frequencies of unfractionated and purified cells with integrated HIV-1 DNA. These variances were calculated from the confidence intervals given in ref. 29. The conservative assumption of independence was made.

§ Calculated using the fraction of CD4⁺DR⁺ cells among lymph node cells assuming a total body population of 10¹² lymphocytes.

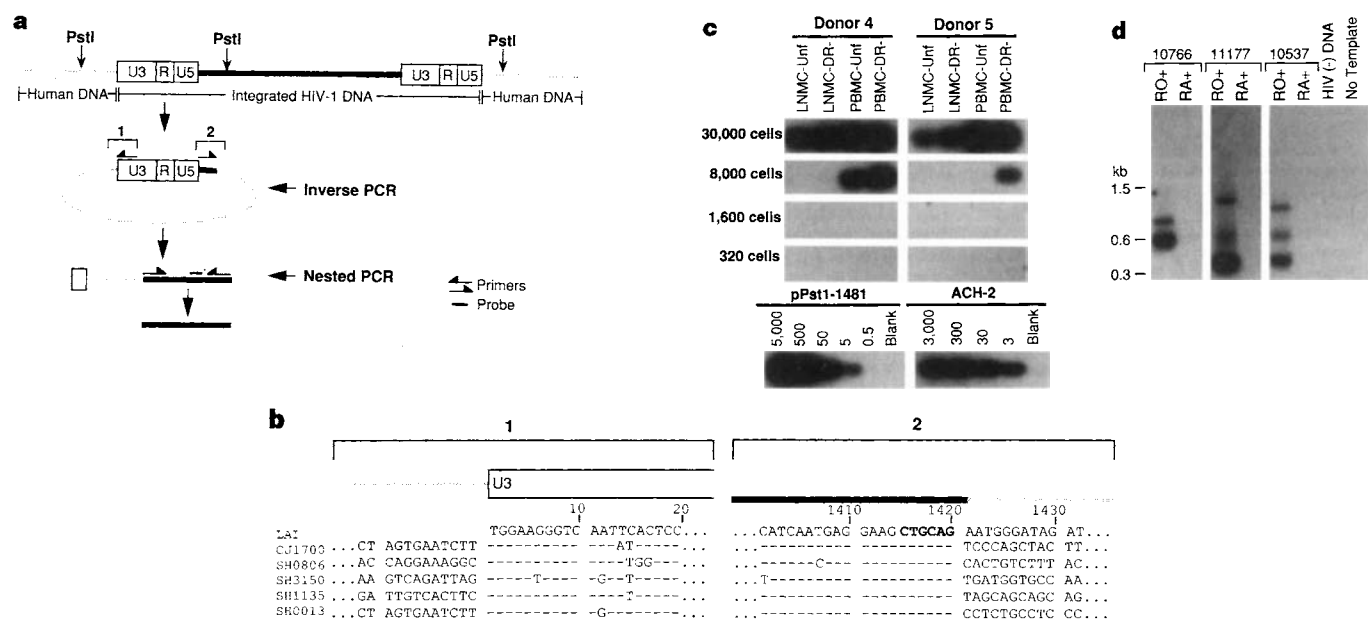


Figure 1 Quantitative analysis of integrated HIV-1 DNA in resting CD4⁺ T cells. **a**, Inverse PCR assay for integrated HIV-1 DNA. DNA from purified resting CD4⁺ T cells was digested with *Pst*I, serially diluted into DNA from an HIV-1-negative donor, and subjected to ligation under dilute conditions followed by inverse PCR amplification using outward-directed LTR and *gag* primers. This amplifies cellular DNA between the integration site and the nearest upstream *Pst*I site. In some experiments, an aliquote of the inverse PCR product was used as a template for a second nested PCR reaction with *gag* primers, and the product was detected by Southern hybridization using an internal probe. '1' and '2' indicate regions of the initial inverse PCR products for which the sequence is shown in Fig. 2b. **b**, Sequences of representative inverse PCR products cloned from five previously characterized²¹ seropositive donors aligned with the reference LAI sequence. **c**,

Representative results of a two-step amplification of integration sites carried out on the indicated number of cell equivalents of donor DNA diluted into 30,000 cell equivalents of DNA from a seronegative donor. Limiting dilution analysis was carried out on unfractionated lymphocyte populations (Unf) and on highly purified resting CD4⁺ T cell populations (DR-) from lymph nodes (LN) and peripheral blood (PB). A plasmid-derived construct mimicking integrated HIV-1 DNA (pPst1-1481) and ACH-2 cells, which carry a single integrated copy of provirus, were diluted into 30,000 cell equivalents of HIV-negative DNA and run in parallel. **d**, Inverse PCR analysis on DNA from 30,000 purified resting peripheral blood CD4⁺ T cells sorted into memory (RO⁺) and naive (RA⁺) subpopulations. Bands representing distinct integration events were detected by Southern blotting following the initial inverse PCR amplification. Donors were from a previously described cohort²¹.

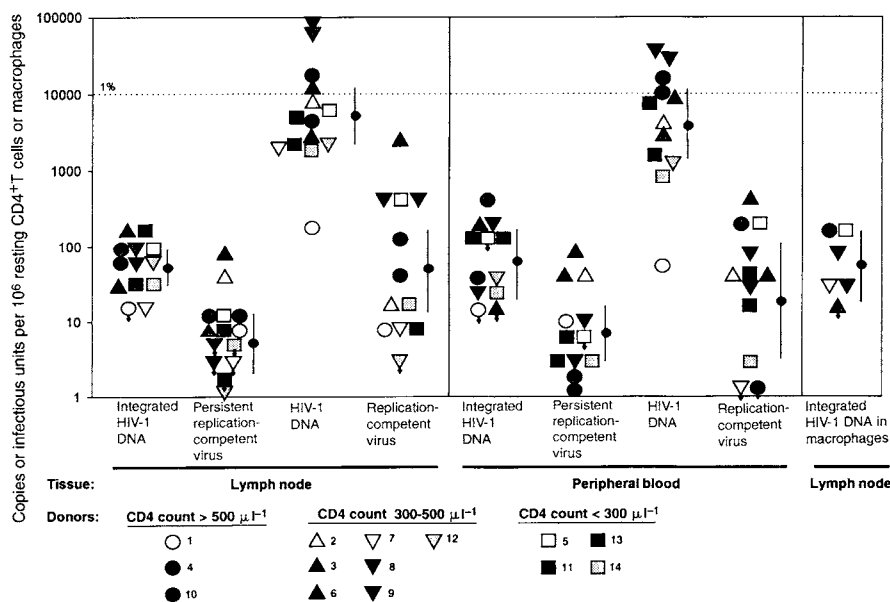


Figure 2 Frequencies of resting CD4⁺ T cells and macrophages carrying various forms of HIV-1. Frequencies are expressed as copies or infectious units per 10⁵ cells. The frequencies of resting CD4⁺ T cells with 'integrated HIV-1 DNA' were determined by inverse PCR analysis as described for Fig. 1. The frequencies of resting CD4⁺ T cells with 'persistent, replication-competent virus' were determined by micrococulture assay of purified resting cells isolated after 6-d preincubation. The frequencies of resting CD4⁺ T cells with 'HIV-1 DNA' were estimated from conventional PCR analysis using 4 different primer pairs (Fig. 3a). The frequencies of cells carrying 'replication-competent virus' were determined by coculture assay without a preincubation period. The frequencies of lymph

node macrophages carrying integrated HIV-1 DNA were determined using the inverse PCR method and are plotted here as 'integrated HIV-1 DNA in lymph node macrophages'. For each assay, mean frequencies and 95% confidence intervals about the mean for the data set are plotted to the right of the individual donor values. In cases where the amount of sample limited analysis at higher cell concentrations, an upper bound is plotted (downward arrow). Measurements of integrated HIV-1 DNA in resting CD4⁺ T cells and macrophages are corrected for the fraction of integration events occurring too far from a *Pst*I site to be detected. Full clinical data on donors are available as Supplementary information.

were done in replicate, the actual limit of detection is even lower (16 copies per 10^6 cells). The measured frequencies of cells with integrated HIV-1 DNA were corrected for integration events occurring >2 kb from *Pst*I sites. Assays of purified resting $CD4^+$ T cells and unfractionated mononuclear cells were generally positive for integrated HIV-1 DNA in the tubes containing 30,000 cell equivalents of DNA (Fig. 1c). However, the signal was undetectable in the tubes containing donor DNA equivalent to 1,600 cells or less. The frequencies of resting $CD4^+$ T cells with integrated HIV-1 DNA, determined from the limiting dilution PCR data, are plotted in Fig. 2. The frequencies ranged from a maximum of 410 per 10^6 resting $CD4^+$ T cells to a minimum of less than 16 per 10^6 cells, the lower limit of detection. Surprisingly, the frequencies were not significantly different in blood and lymph node and did not correlate with CD4 count, plasma RNA or therapy. These findings suggest that in asymptomatic HIV-1 infection, a relatively stable systemic steady state is established in which only a minute fraction ($<0.05\%$) of the resting $CD4^+$ T population carries integrated HIV-1 DNA. The low frequency may reflect the fact that productively infected, activated $CD4^+$ T cells rarely survive both viral cytopathic effects and host cytolytic mechanisms for long enough to revert back to a resting memory state. Infected cells that do survive and revert to a resting state should show a memory phenotype. To test this idea, purified resting $CD4^+$ T cells were subfractionated into naive and memory subsets based on CD45 isoform expression. Inverse PCR bands representing distinct integration events were present only in the memory ($CD45RO^+$) population in 3 of 4 samples examined (Fig. 1d).

To provide an independent confirmation for the low frequency of cells with integrated HIV-1 DNA and to determine what fraction of resting $CD4^+$ T cells can be induced to produce infectious virus, we used quantitative viral culture to detect resting $CD4^+$ T cells carrying replication-competent forms of the virus. Because the resting cell population may contain cells with integrated HIV-1 DNA as well as cells with unintegrated HIV-1 DNA and/or absorbed virions, purified resting $CD4^+$ T cells were preincubated in the absence of activating stimuli for six days to allow the degradation of labile, unintegrated forms of the virus. Serial dilutions of precultured cells were then subjected to activation and co-culture to amplify infectious virus. In control experiments, resting $CD4^+$ T cells from an HIV-1 seronegative donor were infected with HIV-1_{LAI} *in vitro* and activated immediately or after a six-day preincubation period. Infectious virus could be recovered from as few as three infected resting $CD4^+$ T cells activated immediately after infection, demonstrating the sensitivity of the culture system. However, no virus production from precultured cells was observed, although the preculture period did not affect mitogen responsiveness. Thus this assay provides a sensitive means of detecting persistent, replication-competent forms of the virus, including provirus present in an integrated state, and was used to estimate an upper bound on the frequency of cells carrying replication-competent integrated provirus. The minimum number of precultured resting $CD4^+$ T cells required to rescue infectious virus was between 8,000 and $>1,000,000$ for various samples. The calculated frequencies of resting $CD4^+$ T cells carrying persistent replication-competent forms of the virus are plotted in Fig. 2. The mean frequency was 5 per 10^6 cells in lymph nodes and 7 per 10^6 cells in blood. These values provide independent confirmation that the frequency of resting $CD4^+$ T cells in the postintegration state of latency is very low. The frequencies were not significantly different in lymph nodes and blood and were not strongly correlated with CD4 count or plasma viraemia. The maximum frequencies of resting $CD4^+$ T cells with replication-competent integrated provirus, estimated by this culture assay, were lower than the frequencies of resting $CD4^+$ T cells with integrated HIV-1 DNA, as detected by inverse PCR. This suggests that some of the integrated HIV-1 DNA in resting $CD4^+$ T cells is defective.

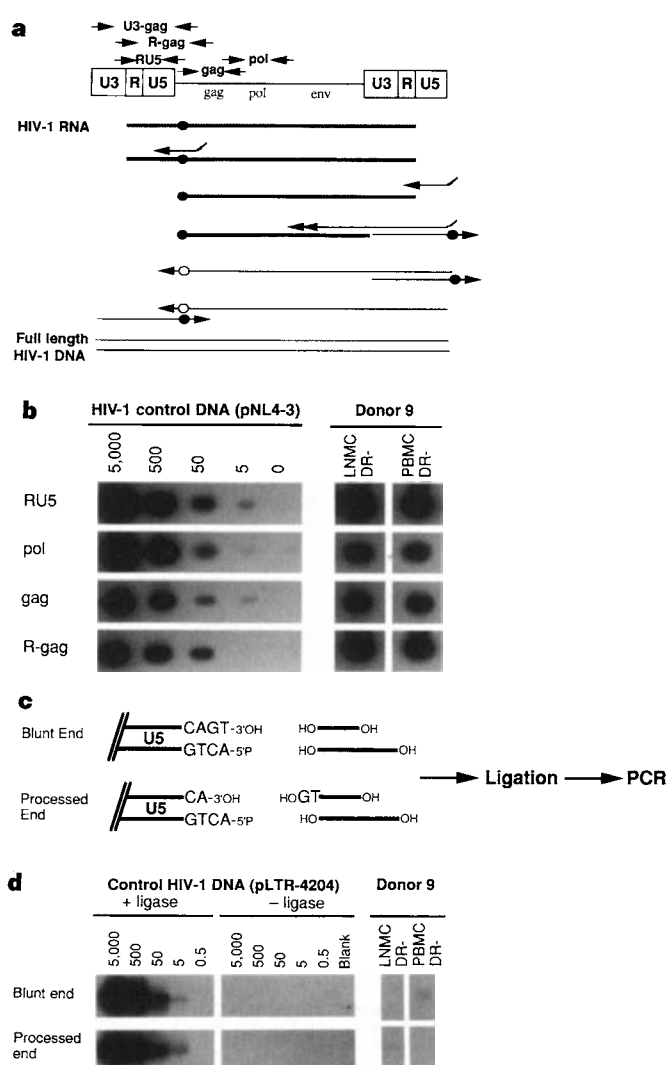


Figure 3 Analysis of HIV-1 DNA in resting $CD4^+$ T cells. **a**, Positions of primer pairs used to detect intermediates in the reverse transcription process. The initial reverse transcription product is detectable with primers in the R and U5 regions of the HIV-1 LTR. Following the first-strand transfer reaction, amplification with primers in *pol* and then *gag* becomes possible. Following the second-strand transfer reaction and polymerization towards the end of each template strand, amplification becomes possible with the LTR-gag and finally the U3-gag primer pairs. **b**, Representative results of amplification of an HIV-1 plasmid standard and donor samples with the indicated primer pairs. Resting (DR-) $CD4^+$ T cells purified from LNMC and PBMC populations of donor 9 showed similar copy numbers of HIV-1 template with all primer pairs. **c**, **d**, Linker ligation assay for blunt-ended and integrase-processed HIV-1 DNA. Full-length, blunt ended molecules were detected by ligation of blunt-ended, asymmetric linkers⁶. To determine the sensitivity of this assay, a plasmid construct carrying the HIV-1 LTR (pLTR-4202) was cut with *Stu*I at the 5' or 3' end of the LTR. The indicated numbers of copies of this plasmid were diluted into 30,000 cell equivalents of DNA from an HIV-1 negative donor. Samples were incubated with a blunt-ended asymmetrical linker in the presence or absence of ligase. Ligation products were amplified with primers hybridizing with LTR and linker sequences and were detected by Southern hybridization. Blunt-ended plasmid-derived standards could be detected down to a level of <5 copies per 30,000 cells. To detect molecules that had been processed by integrase as an initial step in the integration reaction, a parallel set of experiments was carried out using a linker with 5' G-T overhang complementary to the expected integrase cleavage product. In control studies, a plasmid construct carrying the HIV-1 LTR was cleaved at the 5' or 3' end of the LTR using *Acc*I, which leaves a different but comparable dinucleotide overhang. Following ligation with an appropriate complementary linker, amplification and Southern hybridization, these molecules were detectable down to a level of 1 copy per 30,000 cell equivalents of HIV-1-negative DNA.

Using these data and flow measurements of the percentages of resting CD4⁺ T cells in lymph nodes and blood, we estimated the total body number of resting CD4⁺ T cells with integrated HIV-1 DNA. Assuming 10¹² total body lymphocytes, 98% of which are in the lymphoid tissue, the estimated total body load of resting CD4⁺ T cells with integrated HIV-1 DNA ranged from 4.6 × 10⁶ to 3.4 × 10⁷ cells (mean, 1.2 × 10⁷). Using data obtained from the co-culture assay, we also estimated the total number of CD4⁺ T cells carrying persistent, replication-competent forms of the virus. Values ranged from 2.2 × 10⁵ to 1.6 × 10⁷ (mean, 1.4 × 10⁶).

To obtain a more complete picture of viral burden, the frequency of resting CD4⁺ T cells carrying HIV-1 DNA was measured using assays that detect unintegrated as well as integrated forms. Because reverse transcription does not proceed to completion following *in vitro* infection of resting CD4⁺ T cells¹⁹, PCR analysis was carried out using primers that detect different stages of reverse transcription (Fig. 3a). Surprisingly, accumulation of partial reverse transcripts was not observed (Fig. 3b). The number of copies of the RU5 template, which is made during an early step in the reverse transcription process, and the number of copies obtained using primers that detect successively later stages in reverse transcription were generally in the same range (0.5 log units or less), suggesting that most of the reverse transcripts were complete. To confirm these results, additional PCR studies were carried out with a forward primer in the U3 region, very close to the 5' end of the genome and a reverse primer downstream of the primer binding site. In this case, amplification can only occur if the second-strand transfer reaction has taken place and if (–) strand synthesis has proceeded almost to the 5' end of the genome. The number of copies obtained using this primer pair was very close to that obtained with the other primer pairs (not shown). Estimates of the total number of copies of cell-associated HIV-1 DNA, based on the mean values obtained with different primer pairs, are plotted in Fig. 2. The mean frequency of resting CD4⁺ T cells carrying HIV-1 DNA was roughly two logs higher than the frequency observed with a selective assay for integrated HIV-1 DNA, suggesting that most of the cell-associated HIV-1 DNA in resting CD4⁺ T cells *in vivo* is unintegrated. The frequency of resting CD4⁺ T cells with unintegrated HIV-1 DNA was not significantly different in blood and lymph nodes.

To define the state of unintegrated HIV-1 DNA in resting CD4⁺ T cells, we directly assayed for full-length unintegrated proviral HIV-1 DNA using a linker-ligation assay (Fig. 3c). This assay detects double-stranded, blunt-ended DNA molecules and should therefore detect the complete reverse transcript that is the substrate for HIV-1 integrase. A plasmid-derived construct that mimics full-length, blunt-ended, unintegrated HIV-1 DNA could be detected at the level of less than five copies diluted into 30,000 cell equivalents of HIV-1-negative DNA (Fig. 3d). However, in the analysis of seven lymph-node samples, the maximum frequency of resting CD4⁺ T cells with blunt-ended, unintegrated HIV-1 DNA was 170 copies per 10⁶ cells. In most cases, the frequency was <170 copies per 10⁶ cells. Thus most unintegrated HIV-1 DNA molecules in resting CD4⁺ T cells *in vivo* do not have detectable free blunt ends, even though reverse transcription has proceeded largely to completion. Because integrase cleaves a G-T dinucleotide from the 3' end of each strand of full-length linear HIV-1 DNA as an initial step in the integration reaction, and because integrase-processed molecules may accumulate in infected cells²³, assays were also carried out with a linker containing a 5' G-T overhang that would selectively ligate to integrase-processed HIV-1 DNA (Fig. 3c, d). Although plasmid-derived standards with a comparable dinucleotide overhang could be readily detected with single-molecule sensitivity in the presence of 30,000 cell equivalents of DNA, no signal was observed in 11 of 14 donor samples, indicating a frequency of <33 copies per 10⁶ cells. For the remaining three donors, the signal was just above the threshold of detection at 33 copies per 10⁶ cells. Full-length HIV-1 DNA molecules can also undergo circularization reactions. We

directly assayed for one and two LTR circles using PCR assays with a sensitivity of <5 copies in 30,000 genome equivalents of human DNA. Both 1 and 2 LTR circles could be detected, but neither form accounted for more than 10% of the unintegrated HIV-1 DNA in resting CD4⁺ T cells (not shown). Taken together, these results suggest that most HIV-1 DNA detected in resting CD4⁺ T cells is full-length, linear and unintegrated, but lacks blunt ends recognizable by integrase.

To determine the overall frequency of cells carrying replication-competent forms of the virus, we also did a quantitative culture assay on purified resting CD4⁺ T cells in which cells were activated immediately after isolation. This assay detects all replication-competent forms of the virus, including labile and persistent forms (Fig. 2). Values were generally substantially lower than the frequencies of cells carrying HIV-1 DNA as detected by PCR. This difference reflects the fact that only a minor fraction of the unintegrated HIV-1 DNA in resting CD4⁺ T cells is replication-competent, consistent with the molecular assays which detected few molecules with blunt ends that were recognizable by integrase.

Viral load in activated CD4⁺ T cells was estimated by comparing values obtained with resting cells and with unfractionated lymphocyte populations depleted of adherent cells. Viral load in the unfractionated populations, which contain activated as well as resting CD4⁺ T cells, was not generally higher than viral load in purified resting CD4⁺ T cells. By assuming that all of the cell-associated HIV-1 in lymphocytes was either in resting or activated CD4⁺ T cells⁵, the frequency and total body load of activated CD4⁺ T cells with integrated HIV-1 DNA could be estimated (Table 1). The estimated mean frequency of cells with integrated HIV-1 DNA among activated CD4⁺ T cells was 224 per 10⁶ cells (~0.02%) in the lymph nodes. Thus, even for activated CD4⁺ T cells, a preferred target cell for viral replication, only a minute fraction of the cells are stably infected at any given time. The total body load of activated CD4⁺ T cells with integrated provirus was estimated at 3.1 × 10⁷ (range, 1.2 × 10⁷–6.7 × 10⁷). Given that a single productively infected CD4⁺ T cell can produce 10³ virions²⁴, this surprisingly low estimate of the number of activated CD4⁺ T cells with integrated HIV-1 DNA can nevertheless account for a significant part of the observed rate of virion production (~10¹⁰ per day)¹⁷ in infected individuals. The frequency of activated CD4⁺ T cells with HIV-1 DNA detectable with standard PCR was also estimated. These frequencies were dramatically higher than the frequency of activated cells with integrated HIV-1 DNA. The mean frequency was 18,530 per 10⁶ cells or ~1.9% of activated CD4⁺ T cells in lymphoid tissues (range, 0.2–22.6%). This may reflect the fact that the cells in which integration does occur die rapidly as a result of viral cytopathic effects or host effector mechanisms. Alternatively, as already discussed, many of the unintegrated HIV-1 DNA molecules may be unable to undergo integration. This contention is supported by limiting dilution coculture studies showing that the fraction of activated CD4⁺ T cells from which infectious virus can be obtained is much lower than the fraction of activated CD4⁺ T cell with HIV-1 DNA (not shown).

Macrophages can be productively infected by HIV-1 (ref. 25) and may function as a major reservoir for the virus as these cells are relatively resistant to viral cytopathic effects. Therefore, viral load was also examined in lymph node macrophages using inverse PCR. As shown in Fig. 2, the mean frequency of lymph node macrophages with integrated HIV-1 DNA was very low at 54 per 10⁶ macrophages.

Although the present study provides only a 'snapshot' picture of viral load in HIV-1 infection, several conclusions may be drawn. HIV-1 infection persists and progresses even though at any given time only a minute fraction of the susceptible cell populations are stably infected, as assessed by the presence of integrated provirus. This is true for resting and activated CD4⁺ T cells and for macrophages in the lymphoid tissues. By cross-sectional analysis, CD4 depletion is not accompanied by any dramatic increase in the load of

integrated virus in the remaining CD4⁺ T cells. The reservoir of resting CD4⁺ T cells with integrated provirus is <0.05% of the resting CD4⁺ T-cell population. The surprisingly small size of this reservoir may be a reflection of the fact that productively infected cells turn over quickly, due to viral cytopathic effects or host cytotoxic T-lymphocyte (CTL) activity, and therefore rarely survive long enough to revert back to a memory state carrying integrated provirus. The fraction of resting CD4⁺ T cells with integrated provirus is similar in blood and lymph nodes, consistent with the continual trafficking of these cells between various lymphoid compartments. Thus, meaningful studies of this potential latent reservoir can be carried out using the much more easily obtained peripheral blood samples. In attempting to reconcile the low numbers of latently infected cells detected here with results from other studies, it must be kept in mind that most of the viral DNA detected by standard PCR is likely to be unintegrated, as suggested by Bukrinsky *et al.*⁶ and confirmed here. In both activated and resting CD4⁺ T cells, the dominant species of HIV-1 DNA is an unintegrated form that is full-length, or nearly full-length, and linear. Most of these DNA molecules lack the correct blunt ends of the type that are recognized by integrase. Some of the unintegrated HIV-1 DNA is in the form of 1 and 2 LTR circles (<10%). The remainder may be linear molecules with defects at the ends resulting from the errors in reverse transcription or the action of host nucleases. Culture assays indicate that only a small fraction of this unintegrated HIV-1 DNA is replication-competent. The unintegrated HIV-1 DNA may be labile, and thus effective antiretroviral therapy should quickly reduce cell-associated viral load to a level that reflects a more stable, integrated reservoir²⁶. The size of the latent reservoir as measured here in several logs lower than estimates based on PCR that do not distinguish between integrated and unintegrated HIV-1 DNA. However, although latently infected CD4⁺ T cells are present at low frequency, the importance of this small reservoir should not be underestimated because memory CD4⁺ T cells can survive for months and possibly years²⁷. Defining the rate of turnover of this reservoir will be an important next step³⁰. □

Methods

Details are available as Supplementary Information.

Tissue samples. Excisional biopsies of enlarged superficial nodes were obtained from healthy adult volunteers with asymptomatic HIV-1 infection. All donors gave informed consent. Macrophages were isolated by adherence. Resting CD4⁺ cells were isolated as described²¹.

PCR analysis. Integrated HIV-1 DNA was detected using an inverse PCR procedure²¹ with a second nested amplification of an internal *gag* sequences. The method was adapted to a limiting dilution format by serially diluting donor DNA into DNA from an HIV-1-negative individual. Controls in which ligase was omitted or the first PCR was run for zero cycles were invariably negative. Integrity of DNA templates was confirmed by amplification of a β -globin sequence.

Partial reverse transcripts, full-length viral DNA, and 1 and 2 LTR circles were quantified by PCR using appropriate plasmid derived standards. The primer pairs used to detect distinct stages of reverse transcription were based ref. 19, but were modified to include only highly conserved regions of the HIV-1 genome. After Southern blot hybridization, bands were quantified by Phosphor Imager analysis using a standard curve based on known copy numbers of the relevant plasmid-derived control constructs. To determine the frequency of cells carrying HIV-1 DNA, the geometric mean frequency was calculated from values obtained with four different primer pairs. For some donors, copy numbers obtained with a single primer pair were consistently much lower than those obtained with the other three primer pairs, probably owing to minor sequence differences within the highly conserved regions chosen as primer sites. In this case, the outlying low value was not included.

Blunt-ended HIV-1 DNA was detected by linker-ligation⁶. To detect molecules processed by integrase, a similar procedure was followed except that the linker had a 5' G-T overhang. A plasmid, pLTR-4202, with appropriate restriction sites immediately 5' or 3' to the HIV-1 LTR sequence, was used to

generate standards with blunt or recessed ends. These standards were diluted in known copy numbers into HIV-1-negative DNA and run in parallel to allow quantification of donor samples.

Culture studies. To determine the frequency of resting CD4⁺ T cells carrying persistent, replication-competent forms of the virus, highly purified resting cells were cultured for 6 d and then counted and seeded in duplicate 5-fold serial dilutions in interleukin (IL)-2-containing medium. To induce activation of resting CD4⁺ T cells, phytohaemagglutinin (PHA) and irradiated peripheral blood mononuclear cells (PBMC) were added. These conditions allow activation of virtually 100% of resting T cells²⁸. The following day, PHA was removed and 1–3-day PHA-stimulated, CD8-depleted PBMC from an HIV-1 negative donor were added to amplify infectious virus. Cultures were fed and split as needed and additional CD8-depleted PHA blasts were added on day 7. Supernatants collected on day 14 were analysed for p24 antigen by ELISA. To determine the frequency of cells with all replication-competent forms of the virus (labile and persistent), purified resting CD4⁺ T cells were placed in limiting dilution microculture immediately, without 6-day preculture.

Mathematical and statistical methods. Limiting-dilution inverse PCR and virus culture data were analysed by a maximum likelihood method²⁹. Duplicate 5-fold dilutions were set up starting with 10⁶ cells per tube or well. Individual tubes or wells were scored as positive or negative, and the results used to determine infected cell frequencies as described²⁹. The 95% confidence intervals for individual determinations spanned about 1.4 logs. In cases where all determinations were negative, an upper bound on the infected cell frequency was estimated by making the conservative assumption that the next lowest dilution (highest cell concentration) would have been positive.

In the inverse PCR analysis of the frequency of cells carrying integrated HIV-1 DNA, the following procedure was used to correct for the fraction of integration events occurring too far from a *Pst*I site to be detected. The probability of detection, $P(d)$, of an integration event was estimated by:

$$P(d) = \sum_L P(d|L)P(L)$$

in which the $P(d|L)$ is the probability of detecting an integration event occurring within a *Pst*I fragment of length L , and $P(L)$ is the probability of a *Pst*I fragment being of length L . $P(d|L)$ was determined by carrying out the entire digestion, ligation and inverse PCR analysis on 200 ng HIV-1-negative DNA to which had been added known copy numbers of plasmid constructs that gave inverse PCR products of different sizes. $P(L)$ was determined from the weight-average distribution of *Pst*I fragments of human genomic DNA. The true frequency of cells with integrated provirus was estimated by dividing the measured frequencies by $P(d)$ to account for underdetection. The value of $P(d)$ was 0.5, based on analysis of genomic DNA from 3 different donors.

Received 6 January; accepted 4 April 1997.

- Ho, D. D., Moudgil, T. & Alam, M. Quantitation of human immunodeficiency virus type 1 in the blood of infected persons. *N. Engl. J. Med.* **321**, 1621–1625 (1989).
- Coombs, R. W. *et al.* Plasma viremia in human immunodeficiency virus infection. *N. Engl. J. Med.* **321**, 1626–1631 (1989).
- Piatk, M. Jr. *et al.* High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR. *Science* **259**, 1749–1754 (1993).
- Mellors, J. W. *et al.* Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science* **272**, 1167–1170 (1986).
- Schnittman, S. M. *et al.* The reservoir for HIV-1 in human peripheral blood is a T cell that maintains expression of CD4. *Science* **245**, 305–308 (1989).
- Bukrinsky, M. I., Stanwick, T. L., Dempsey, M. P. & Stevenson, M. Quiescent T lymphocytes as an inducible virus reservoir in HIV-1 infection. *Science* **254**, 423–427 (1991).
- Bagasra, O., Hauptman, S. P., Lischner, H. W., Sachs, M. & Pomerantz, R. J. Detection of human immunodeficiency virus type 1 provirus in mononuclear cells by *in situ* polymerase chain reaction. *N. Engl. J. Med.* **326**, 1385–1391 (1992).
- Patterson, B. K. *et al.* Detection of HIV-1 DNA and messenger RNA in individual cells by PCR-driven *in situ* hybridization and flow cytometry. *Science* **260**, 976–979 (1993).
- Embreton, J. *et al.* Massive covert infection of helper T lymphocytes and macrophages by HIV during the incubation period of AIDS. *Nature* **362**, 359–362 (1993).
- Pantaleo, G. *et al.* HIV infection is active and progressive in lymphoid tissue during the clinically latent stage of disease. *Nature* **362**, 355–358 (1993).
- Harper, M. E., Marselle, L. M., Gallo, R. C. & Wong-Staal, F. Detection of lymphocytes expressing human T-lymphotropic virus type III in lymph nodes and peripheral blood from infected individuals by *in situ* hybridization. *Proc. Natl Acad. Sci. USA* **83**, 772–776 (1986).
- Seshamma, T., Bagasra, O., Trono, D., Baltimore, D. & Pomerantz, R. J. Blocked early-stage latency in the peripheral blood cells of certain individuals infected with human immunodeficiency virus type 1. *Proc. Natl Acad. Sci. USA* **89**, 10663–10667 (1992).
- Adams, M. *et al.* Cellular latency in human immunodeficiency virus-infected individuals with high CD4 levels can be detected by the presence of promoter-proximal transcripts. *Proc. Natl Acad. Sci. USA* **91**, 3862–3866 (1994).
- Haase, A. T. *et al.* Quantitative image analysis of HIV-1 infection in lymphoid tissue. *Science* **274**, 985–989 (1996).

15. Ho, D. D. *et al.* Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* **373**, 123–126 (1995).
16. Wei, X. *et al.* Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* **373**, 117–122 (1995).
17. Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M. & Ho, D. D. HIV-1 dynamics *in vivo*: virion clearance rate, infected cell life-span, viral generation time. *Science* **271**, 1582–1586 (1986).
18. Coffin, J. M. HIV population dynamics *in vivo*: Implications for genetic variation, pathogenesis, and therapy. *Science* **267**, 483–489 (1995).
19. Zack, J. A. *et al.* HIV-1 entry into quiescent primary lymphocytes: molecular analysis reveals a labile, latent viral structure. *Cell* **61**, 213–222 (1990).
20. Bukrinsky, M. I. *et al.* Active nuclear import of human immunodeficiency virus type 1 preintegration complexes. *Proc. Natl Acad. Sci. USA* **89**, 6580–6584 (1992).
21. Chun, T.-W. *et al.* Fate of HIV-1-infected T cells *in vivo*: Rates of transition to stable latency. *Nature Med.* **1**, 1284–1290 (1995).
22. Pauza, C. D. Two bases are deleted from the termini of HIV-1 linear DNA during integrative recombination. *Virology* **179**, 886–889 (1990).
23. Brown, P. O., Bowerman, B., Varmus, H. E. & Bishop, J. M. Retroviral integration: Structure of the initial covalent product and its precursor, and a role for the viral IN protein. *Proc. Natl Acad. Sci. USA* **86**, 2525–2529 (1989).
24. Dimitrov, D. S. *et al.* Quantitation of human immunodeficiency virus type 1 infection kinetics. *J. Virol.* **67**, 2182–2190 (1993).
25. Gartner, S. *et al.* The role of mononuclear phagocytes in HTLV-III/LAV infection. *Science* **233**, 215–219 (1986).
26. Dickover, R. E. *et al.* Decreases in unintegrated HIV DNA are associated with antiretroviral therapy in AIDS patients. *J. Acq. Imm. Def. Syndr.* **5**, 31–36 (1992).
27. Michie, C. A., McLean, A., Alcock, C. & Beverley, P. C. Lifespan of human lymphocyte subsets defined by CD45 isoforms. *Nature* **360**, 264–265 (1992).
28. Patel, S. S., Duby, A. D., Tiele, D. L. & Lipsky, P. E. Phenotypic and functional characterization of human T cell clones. *J. Immunol.* **141**, 3726–3736 (1988).
29. Myers, L. A., McQuay, L. J. & Hollinger, F. B. Dilution assay statistics. *J. Clin. Microbiol.* **32**, 732–739 (1994).
30. Perelson, A. S. *et al.* Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature* **387**, 188–191 (1997).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from Mary Sheehan at the London editorial office of *Nature*.

Acknowledgements. We thank E. Cornell for coordinating clinical aspects of the study, P. Lietman, J. Boeke, R. Markum and M. Schlissel for discussion and A. Siliciano for editorial assistance. This study was supported by the NIH and by the Johns Hopkins Outpatient General Clinical Research Center.

Correspondence and requests for materials should be addressed to R.E.S. (e-mail: rsiliciano@welchlink.welch.jhu.edu).

Decay characteristics of HIV-1-infected compartments during combination therapy

Alan S. Perelson, Paulina Essunger, Yunzhen Cao*, Mika Vesanen*, Arlene Hurley*, Kalle Saksela†, Martin Markowitz* & David D. Ho*

Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

* Aaron Diamond AIDS Research Center, The Rockefeller University, 455 First Avenue, New York, New York 10016, USA

† Institute of Medical Technology, University of Tampere, Tampere, Finland

Analysis of changes in viral load after initiation of treatment with potent antiretroviral agents has provided substantial insight into the dynamics of human immunodeficiency virus type 1 (HIV-1)^{1–3}. The concentration of HIV-1 in plasma drops by ~99% in the first two weeks of treatment owing to the rapid elimination of free virus with a half-life ($t_{1/2}$) of ≤ 6 hours and loss of productively infected cells with a $t_{1/2}$ of 1.6 days³. Here we show that with combination therapy this initial decrease is followed by a slower second-phase decay of plasma viraemia. Detailed mathematical analysis shows that the loss of long-lived infected cells ($t_{1/2}$ of 1–4 weeks) is a major contributor to the second phase, whereas the activation of latently infected lymphocytes ($t_{1/2}$ of 0.5–2 weeks) is only a minor source. Based on these decay characteristics, we estimate that 2.3–3.1 years of a completely inhibitory treatment would be required to eliminate HIV-1 from these compartments. To eradicate HIV-1 completely, even longer treatment may be needed because of the possible existence of undetected viral compartments or sanctuary sites.

Eight HIV-1-infected patients were studied who were naive to antiretroviral agents, with baseline CD4 lymphocyte counts ranging from 26 to 505 per μl (Table 1). *De novo* HIV-1 replication in the patients was blocked by the combined use of a protease inhibitor, nelfinavir⁴ (2,250 mg d^{-1}), and two reverse transcriptase inhibitors, zidovudine (600 mg d^{-1}) and lamivudine (300 mg d^{-1}). After treatment was started, the HIV-1 RNA concentration in plasma was measured weekly for the first month and then every two weeks using the branched DNA (bDNA) assay^{2,3}. Each patient responded with a similar pattern of viral decay (Fig. 1): an initial rapid exponential decline of nearly 2-logs (first phase), consistent with previous observations^{1–3}, followed by a slower exponential decline (second phase). The antiretroviral effect was potent in that plasma viraemia in all eight patients dropped below the standard detection threshold of 500 copies ml^{-1} by 8 weeks of treatment, and was found to be <25 copies ml^{-1} at week 16–20 by using an ultra-sensitive modification of the bDNA assay (data not shown). In addition, no infectious HIV-1 was detectable in 10^7 peripheral blood mononuclear cells (PBMC) from each of the patients after 8 weeks of treatment. These findings not only attest to the potency of the antiretroviral regimen, but also to the lack of emergence of drug-resistant virus during the study period.

The second phase in the decay profile is probably due to sources of HIV-1 not included in our previous analysis³, such as infected tissue macrophages or dendritic cells, activation of latently infected lymphocytes, or release of trapped virions^{2,3,5,6}. If such 'secondary' sources are initially responsible for, say, 1% of the virions in plasma, then the rapid decline in viraemia would slow when the productively infected CD4⁺ T cells, T^* , had decayed to about 1% of their initial level. Further decay of T^* at their fast $t_{1/2}$ would then leave the secondary sources as the major producers of virions. The slower decay of the major secondary source would then be reflected in the slope of the second-phase decline.

Here we incorporate secondary sources of virus into a new model and analyse data obtained from the patients. The model contains cells, M , which upon infection with a rate constant, k_M , become long-lived infected cells, M^* , which produce virus continuously at an average rate per cell, p , and are lost with a rate constant μ_M . We also assume that when CD4⁺ T cells, T , are infected, T^* cells are generated with a rate constant k and that latently infected T cells containing infectious provirus are generated with a rate constant fk , smaller by a factor f . Latently infected lymphocytes, L , are assumed to die with a rate constant δ_L , and to be activated into productively infected cells with rate constant a , giving a total rate constant of loss $\mu_L = a + \delta_L$. T^* are lost with a rate constant, δ , and are assumed to produce a total of N virions (the burst size) during their lifetime. Virions, V , are cleared with a rate constant c . Thus, as a new kinetic model, we propose: $dT^*/dt = kVT + aL - \delta T^*$; $dL/dt = fkVT - \mu_L L$; $dM^*/dt = k_M VM - \mu_M M^*$ and $dV/dt = N\delta T^* + pM^* - cV$.

Virions are produced by infected cells in both tissues and blood, although most such cells will be in tissues⁷. Using methods similar to those already published³, assuming that both HIV-1 reverse transcriptase and protease are completely inhibited by the antiretroviral regimen and that the system was at steady-state before treatment, with baseline viral load V_0 and CD4⁺ T-cell count T_0 , the level of plasma virus after drug therapy should decay as follows:

$$V(t) = V_0[Ae^{-\delta t} + Be^{-\mu_L t} + Ce^{-\mu_M t} + (1 - A - B - C)e^{-ct}] \quad (1)$$

where

$$A = \frac{NkT_0}{c - \delta} \left(1 - \frac{af}{\delta - \mu_L} \right)$$

$$B = \frac{af\delta NkT_0}{\mu_L(\delta - \mu_L)(c - \mu_L)}$$

$$C = \frac{c - NkT_0 \left(1 + \frac{af}{\mu_L}\right)}{c - \mu_M}$$

Here pharmacokinetic and intracellular delays⁸ are ignored as we are interested in effects that occur on a timescale of weeks to months. If the drugs are not completely inhibitory, then the model needs to be supplemented by equations for the uninfected populations T and M . When $f = 0$, no latently infected T cells are generated and this solution reduces to one for a model with the second phase being generated solely by long-lived cells. We call this the long-lived infected cell model.

We observed, down to the HIV-1 RNA detection limit, only two

phases in the viral decay curves of each of the eight patients. This suggests that in each patient there is only one major secondary source of virions. Fitting the plasma viraemia data with the long-lived infected cell model, using nonlinear least-squares regression, we estimate the parameters δ , μ_M , and a composite parameter, NkT_0 . Virion clearance occurs too rapidly to estimate c from the available data in this study. Thus, c was held constant at the mean value, $c = 3 \text{ d}^{-1}$, determined previously³. Three examples of the best-fit curves generated from equation (1), with $f = 0$, are shown in Fig. 1.

The equation for $V(t)$ can also be reduced to the solution of a model with only latently infected T cells instead of long-lived cells. Fitting the plasma viraemia data with this latently-infected cell model gave fits that were indistinguishable from that of the long-lived

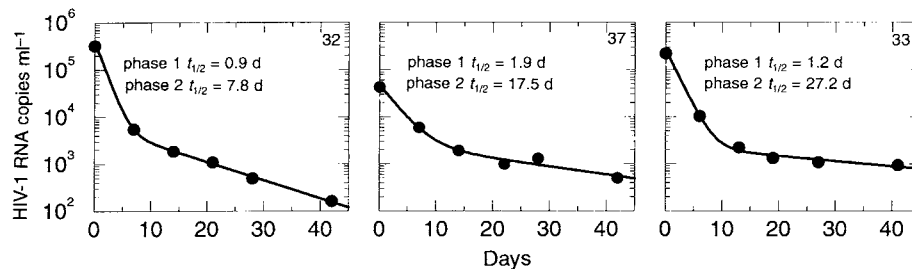


Figure 1 Plasma levels of HIV-1 RNA (filled circles) measured by the bDNA assay for three representative patients during triple therapy, begun on day 0; data points between 100 and 500 copies ml^{-1} were determined using 10 ml plasma. The theoretical curve (solid line) was obtained by nonlinear least-square fitting of the

logarithm of equation (1), with $f = 0$, to the logarithm of the data. The parameters δ , μ_M and NkT_0 , were simultaneously estimated. The parameter estimates obtained were: patient 32, see Table 1; patient 33, $\delta = 0.57 \text{ d}^{-1}$, $\mu_M = 0.025 \text{ d}^{-1}$, and $NkT_0 = 2.97$; patient 37, $\delta = 0.36 \text{ d}^{-1}$, $\mu_M = 0.040 \text{ d}^{-1}$, and $NkT_0 = 2.80$.

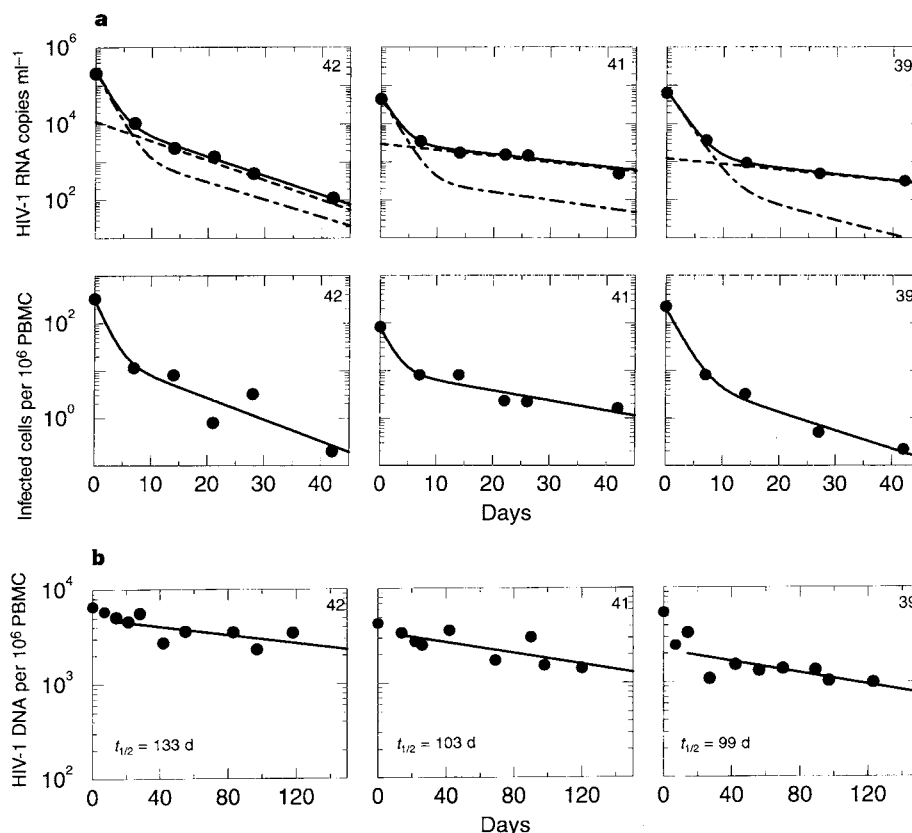


Figure 2 a, Plasma levels of HIV-1 RNA, PBMC infectivity, and **b**, HIV-1 proviral DNA for three representative patients. The theoretical curves (solid lines) were obtained by fitting equations (1) and (2) simultaneously to the data (filled circles). The long-dashed short-dashed line indicates the contribution to the total HIV-1 RNA by pre-existing productively infected T cells plus those generated by the gradual activation of latently infected T cells (the terms $Ae^{-\delta t} + Be^{-\mu_L t}$ in equation (1)); the dashed line indicates the contribution made by long-lived infected cells,

M^* (the term $Ce^{-\mu_M t}$ in equation (1)). By day 14, the productively infected CD4^+ T cells initially present have substantially decayed, and the remaining productively infected T cells are being generated by activation of latently infected T cells. As indicated by the long-dashed short-dashed line after day 14, the contribution to the second phase made by latently infected T cells is substantially smaller than the contribution made by long-lived infected cells (dashed line).

infected cell model for each of the eight patients. Thus, additional viral-load data were obtained in order to distinguish between the models. At each time point, the infectivity titre of HIV-1 in PBMC was measured by limiting dilution (5-fold serially; quadruplicate samples), as described^{9,10}. The number of infected cells per 10⁶ PBMC was estimated using maximum likelihood¹¹. In patient 40, infectivity dropped to below the detection threshold too rapidly to permit a detailed analysis. Patient 32 has a plasma viral load that was not in steady state before treatment and therefore could not be analysed further.

In the PBMC infectivity assays, cells that latently harbour infectious provirus would be activated *in vitro* and thus would be scored in the limiting-dilution measurement. If we assume that long-lived, virus-producing cells reside in tissues and that the number of infected CD4⁺ T cells in blood are proportional to the total number in the body, then the number of PBMC carrying infectious HIV-1 detected in the PBMC infectivity assay is proportional to $I(t) = T^*(t) + L(t)$, where

$$I(t) = \frac{kV_0T_0}{\delta} \left[\left(1 - \frac{af}{\delta - \mu_L} \right) e^{-\delta t} + \frac{f\delta}{\mu_L} \left(1 + \frac{a}{\delta - \mu_L} \right) e^{-\mu_L t} \right] \quad (2)$$

Simultaneously fitting equations (1) and (2) to the plasma viraemia and PBMC infectivity data yielded estimates of the various decay rate constants (Table 1). In general, 10–12 data points were used in the fitting, with an average of 6.4 data points specifically during the second phase (see Fig. 2a for examples). The parameter, μ_M , for long-lived infected cells ranged from 0.028 to 0.119 d⁻¹, with a mean value of 0.061 ± 0.033 d⁻¹. The corresponding half-life, $t_{1/2} = \ln 2/\mu_M$, had a mean value of 14.1 ± 7.5 d, and was negatively correlated with the logarithm of the baseline viral load ($r^2 = 0.77$, $p < 0.03$). The parameter, δ , for productively infected T cells ranged from 0.36 to 1.2 d⁻¹, with a mean of 0.69 ± 0.25 d⁻¹. The corresponding $t_{1/2}$ values ranged from 0.6 to 1.9 d, with a mean of 1.1 ± 0.4 d. The parameter, μ_L , for cells latently carrying infectious provirus ranged from 0.049 to 0.181 d⁻¹, with a mean of 0.098 ± 0.048 d⁻¹. The corresponding $t_{1/2}$ ranged from 3.8 to 14.1 d, with a mean of 8.5 ± 4.0 d. As $c = 3$ d⁻¹ was a minimal estimate³, we also examined the effect of using $c = 30$ d⁻¹. This led to minor changes in our parameter estimates (<10% for δ , <5% for μ_M and μ_L).

Most infected PBMC harbour defective provirus^{1,12} and cannot be activated into virus production. We measured the rate of decay of PBMC proviral DNA, δ_{DNA} , by a standard method¹³. In several patients, the DNA data suggested a more rapid first phase, possibly due to the loss of transiently and productively infected cells. Here we are concerned with the turnover of stably infected cells and only

Table 2 Length of treatment (yr) to eradicate HIV-1 from long-lived cells

Second phase half-life	Initial number of infected long-lived cells: 10 ¹²	
	10 ⁹	10 ¹²
1 week	0.6	0.8
2 weeks	1.2	1.5
4 weeks	2.3	3.1

analyse the data from day 14 onwards. As shown in Fig. 2b and Table 1, δ_{DNA} ranged between 0.0016 and 0.033 d⁻¹, with a mean of 0.012 ± 0.011 d⁻¹. The corresponding $t_{1/2}$ values ranged from 21 to 433 d, with a mean of 145 d, which is somewhat longer than previous estimates^{1,12}, but similar to the $t_{1/2}$ of CD4⁺ CD45RO⁺ T cells¹⁴. The rate of proviral decay was therefore substantially slower than that of plasma viraemia or PBMC infectivity decay, suggesting that this DNA is mainly harboured within cells that contain proviruses that cannot be activated to produce infectious virus. While latently infected T cells are a subset of the provirus-containing cells, we chose to estimate δ_L by δ_{DNA} , because in fitting equations (1) and (2), we could not separately determine the loss parameters a and δ_L . Changing δ_L tenfold led only to minor changes (<10%) in the $t_{1/2}$ estimates shown in Table 1.

From the parameter estimates and equation (1) for $V(t)$, we can compute the contribution of latently infected T cells and long-lived cells to the viral load. As shown in Fig. 2a, long-lived infected cells are the major contributor of virions to the second phase of plasma viraemia decay. Furthermore, at the pretreatment steady state, the fraction of plasma virions produced by long-lived infected cells ranged from 1 to 7% (Table 1), whereas the contribution by activation of latently infected T cells was $\leq 1\%$ (Fig. 2a). Thus, 93–99% of the plasma viraemia at steady state was sustained by productively infected T cells, as we reported previously³. From the steady-state conditions and parameter values in Table 1, we computed that the ratio T_0^*/L_0 had a mean of 15.9 ± 11.3 , which is similar to the value previously reported^{15,16}. Latently infected T cells with stably integrated DNA may arise from productively infected CD4⁺ T cells reverting to a resting state¹⁶. Modifying our model so that L arises from T^* rather than from T , we find little difference in our estimate of $t_{1/2}$ for latently infected T cells or T_0^*/L_0 (data not shown).

We find that the principal contributor to the second phase of plasma viral decay has a $t_{1/2}$ of 5.8 to 24.8 d (Table 1). Based on experimental observations that some tissue macrophages are infected *in vivo*^{17–19} and that macrophages infected *in vitro* can continuously release virions for weeks^{20,21}, the long-lived population defined here may consist of tissue macrophages. Theoretically, virions trapped on follicular dendritic cells in lymphoid tissues

Table 1 Summary of infected-cell-loss rates for eight patients

Parameter estimates from fit of $V(t)$ to plasma viral load and $I(t)$ to cell infectivity data																
Infected cell loss																
Patient	Baseline values		Short-lived cells (T^*)		Long-lived cells (M^*)				Latently infected cells (L)				Proviral DNA		Steady-state values	
					Confidence interval				Confidence interval				δ_{DNA} (d^{-1})	$t_{1/2}$ (d)		
	CD4 cells (per μl)	HIV-1 RNA (per ml)	δ (d^{-1})	$t_{1/2}$ (d)	μ_M (d^{-1})	Lower	Upper	$t_{1/2}$ (d)	μ_L (d^{-1})	Lower	Upper	$t_{1/2}$ (d)			T_0^*/L_0	pM_0^*/cV_0
31	277	95,670	0.36	1.9	0.071	0.016	0.100	9.8	0.111	0.066	0.141	6.2	0.013	53	12.8	0.02
32	58	321,400	0.76	0.9	0.089	0.086	0.092	7.8	NA	NA	NA	NA	0.0016	433	NA	NA
33	505	218,800	0.74	0.9	0.036	0.023	0.052	19.3	0.055	0.028	0.077	12.6	0.025	28	26.4	0.01
37	205	42,940	1.20	0.6	0.028	0.015	0.039	24.8	0.181	0.169	0.189	3.8	0.033	21	0.5	0.04
39	216	65,520	0.51	1.4	0.034	0.030	0.039	20.4	0.088	0.073	0.101	7.9	0.0070	99	31.3	0.02
40	26	648,200	0.81	0.9	0.115	0.115	0.115	6.0	NA	NA	NA	NA	0.0024	289	NA	NA
41	321	45,840	0.58	1.2	0.037	0.031	0.043	18.7	0.049	0.036	0.060	14.1	0.0067	103	8.0	0.07
42	86	208,800	0.63	1.1	0.119	0.106	0.130	5.8	0.105	0.071	0.130	6.6	0.0052	133	16.4	0.06
Mean	212	205,896	0.70	1.1	0.066			14.1	0.098			8.5	0.012	145	15.9	0.04
s.d.	158.7	204,652	0.25	0.4	0.038			7.5	0.048			4.0	0.011	144	11.4	0.02

Baseline HIV-1 RNA values are averages of measurements taken at 3 time points over the month preceding start of therapy; baseline CD4 counts are an average of two measurements. Lower and upper 68% confidence intervals were calculated by a bootstrap method²² in which each experiment was simulated 1,000 times. From the parameter estimates and formulae for the pretreatment quasi-steady state, values of T_0^*/L_0 and pM_0^*/cV_0 were obtained. At steady state, $cV_0 = pM_0^* + N\delta T_0^*$. Thus, pM_0^*/cV_0 is the fraction of virions produced by infected long-lived cells. In our analysis, the parameter c was set to $c = 3$ d⁻¹, δ_L was equal to δ_{DNA} , V_0 and T_0 were set to their baseline values, and the remaining parameters were estimated by nonlinear least-squares regression. Patients 32 and 40 were only analysed using the long-lived cell model. For patient 40, a jackknife fit was done and the most outlying data point ignored. As patient 32 was not in steady state before treatment, the data were fitted to a model in which the coefficients A and C in equation (1) were chosen as arbitrary constants rather than as functions of the steady-state parameters. NA, not applicable.

may be released into the circulation, thus serving as an additional source of viral particles during the second phase. Here, what we have termed long-lived, virus-producing cells could be interpreted to include a tissue compartment releasing virions into plasma with a rate constant p , while decaying with a net rate constant μ_M .

Our results provide evidence that the productively infected cells in blood are representative of the total body population. This follows from the observation that the $t_{1/2}$ of the first phase, determined from the rate of fall of plasma virus, corresponds to the $t_{1/2}$ seen in the initial decline of PBMC infectivity (Fig. 2a). Thus, the rate of decay of productively infected cells in blood reflects the rate of decline of such cells body-wide, suggesting that the levels of infected CD4 cells in blood and tissue are proportional. This conclusion is in good agreement with recent experimental findings²². Further, the rapid initial fall of PBMC infectivity provides the first direct evidence for the rapid decay of productively infected cells, previously inferred from plasma virus decay kinetics³.

Our results have direct implications for the possibility of eradicating HIV-1 from an infected person. If *de novo* infection is completely inhibited by antiretroviral agents, the treatment duration necessary to permit all viral compartments to burn out would depend on the decay $t_{1/2}$ and pool size of each element. Given their rapid kinetics, even as many as 10^{12} pre-existing cell-free virions or 10^{12} productively infected CD4⁺ T cells would be eliminated in less than two months. In this study, we found that the $t_{1/2}$ of infected long-lived cells ranged from about 1 to 4 weeks (Table 1); the $t_{1/2}$ of lymphocytes latently infected with infectious provirus ranged from about 1/2 to 2 weeks. The pool sizes of these two populations probably could not exceed 10^{12} cells, the total number of lymphocytes in an average person²³, or $1-3 \times 10^{11}$, the total number of macrophages²⁴. Available experimental data suggest that in general the pool sizes for M' and L are $\leq 10^9$ and 10^8 , respectively^{7,13,16-20}. Calculations based on the above estimates (Table 2) suggest that these known compartments of virus could be completely eliminated after 2.3–3.1 years of treatment with a 100%-inhibitory antiretroviral regimen. However, to eradicate HIV-1 completely from an infected person, treatment may need to be administered for a longer period, because of the possible existence of small, undetected viral compartments that decay more slowly than the second phase, or sanctuary sites (such as the brain) which are impenetrable by some or all of the antiretroviral drugs. In addition, as can be deduced from the slow decay shown in Fig. 2b, the viral DNA within mononuclear cells that could not be experimentally activated to produce infectious virus will still be present for many years and perhaps for life, leaving open a remote chance that infectious progeny could be produced. These concerns must be properly addressed in future studies. In addition, given the high cost and toxicity of prolonged combination therapy, new strategies must be developed to facilitate the extinction of those viral 'embers' that could rekindle the infection. Although significant progress has been made in the past year in the treatment of HIV-1-infection, it would be wrong to believe that we are close to a cure for AIDS. However, the recent advances in treatment and pathogenesis do warrant a close examination of the feasibility of eradicating HIV-1 from an infected person. □

Received 7 January; accepted 7 April 1997.

1. Wei, X. *et al.* Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* **273**, 117–122 (1995).
2. Ho, D. D. *et al.* Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* **373**, 123–126 (1995).
3. Perelson, A. S. *et al.* HIV-1 dynamics *in vivo*: virion clearance rate, infected cell life-span, and viral generation time. *Science* **271**, 1582–1586 (1996).
4. Patlak, A. K. *et al.* Antiviral and resistance studies of AG1343, an orally bioavailable inhibitor of human immunodeficiency virus protease. *Antimicrob. Agents Chemother.* **40**, 292–297 (1996).
5. Heath, S. L. *et al.* Follicular dendritic cells and human immunodeficiency virus infectivity. *Nature* **377**, 740–744 (1995).
6. Pope, M. *et al.* Conjugates of dendritic cells and memory T lymphocytes from skin facilitate productive infection with HIV-1. *Cell* **78**, 389–398 (1994).
7. Haase, A. T. *et al.* Quantitative image analysis of HIV-1 infection in lymphoid tissue. *Science* **274**, 985–989 (1996).
8. Herz, A. V. *et al.* Viral dynamics *in vivo*: limitations on estimates of intracellular delay and virus decay. *Proc. Natl Acad. Sci. USA* **93**, 7247–7251 (1996).

9. Ho, D. D., Moudgil, T. & Alam, M. Quantitation of human immunodeficiency virus type 1 in the blood of infected persons. *N. Engl. J. Med.* **321**, 1621–1625 (1989).
10. Connor, R. I., Mohri, H., Cao, Y. & Ho, D. D. Increased viral burden and cytopathicity correlate temporally with CD4+ T-lymphocyte decline and clinical progression in human immunodeficiency virus type 1-infected individuals. *J. Virol.* **67**, 1772–1777 (1993).
11. Fazekas de St Groth, S. The evaluation of limiting dilution assays. *J. Immunol. Meth.* **49**, R11–R23 (1982).
12. Nowak, M. A., Bonhoeffer, S., Shaw, G. M. & May, R. M. Antiviral drug treatment: dynamics of resistance in free virus and infected cell populations. *J. Theor. Biol.* **184**, 203–217 (1997).
13. Vesanen, M. *et al.* Stability in controlling viral replication identifies long-term nonprogressors as a distinct subgroup among human immunodeficiency virus type 1-infected persons. *J. Virol.* **70**, 9035–9040 (1996).
14. Mitchie, C. A. *et al.* Lifespan of human lymphocyte subsets defined by CD45 isoforms. *Nature* **360**, 264–265 (1992).
15. Shaw, G. M. Viral and cellular dynamics in HIV-1 infection. *Dixième Colloque des Cent Gardes* (eds Girard, M. & Dodet, B.) 9–11 (Elsevier, New York, 1995).
16. Chun, T.-W. *et al.* *In vivo* fate of HIV-1-infected T cells: quantitative analysis of the transition to stable latency. *Nature Med.* **1**, 1284–1290 (1995).
17. Moreno, P. *et al.* Alveolar macrophages are not an important source of viral production in HIV-1 infected patients. *AIDS* **10**, 682–684 (1996).
18. Koenig, S. *et al.* Detection of AIDS virus in macrophages in brain tissue from AIDS patients with encephalopathy. *Science* **233**, 1089–1093 (1986).
19. Cao, Y. *et al.* Identification and quantitation of HIV-1 in the liver of patients with AIDS. *AIDS* **6**, 65–70 (1992).
20. Ho, D. D., Rota, T. R., Hirsch, M. S. Infection of monocyte-macrophages by human T lymphotropic virus type III. *J. Clin. Invest.* **77**, 1712–1715 (1986).
21. Gartner, S. *et al.* The role of mononuclear phagocytes in HTLV-II/LAV infection. *Science* **233**, 215–219 (1986).
22. Chun, T.-W. *et al.* Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* **387**, 183–188 (1997).
23. Trepel, F. Number and distribution of lymphocytes in man. A critical analysis. *Klin. Wschr.* **52**, 511–515 (1974).
24. van Furth, R. *Inflammation: Basic Principles and Clinical Correlates* (eds Gallin, J. I., Goldstein, I. M. & Snyderman, S.) 325–339 (Raven, New York, 1992).
25. Efron, B. & Tibshirani, R. Bootstrap measures for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.* **1**, 54–77 (1986).

Acknowledgements. We thank the patients for their participation, Agouron Pharmaceutical for supporting the clinical trial, D. Chernoff and colleagues at Chiron for technical assistance, C. Macken for help in analysing the limiting dilution data, and B. Goldstein and B. Sulzer for the use of their nonlinear least-squares fitting package. Part of this work was performed under the auspices of the US Department of Energy. This work was supported by the Rockefeller University General Clinical Research Center, the Aaron Diamond Foundation, the Santa Fe Institute, the Joseph P. Sullivan and Jeanne M. Sullivan Foundation, and by grants from the NIH.

Correspondence and requests for materials should be addressed to D.D.H.

Segmental regulation of *Hoxb-3* by *kreisler*

Miguel Manzanera*, Sabine Cordes††, Chung-Tin Kwan§, Mai Har Sham§, Greg S. Barsh† & Robb Krumlauf*

* Division of Developmental Neurobiology, MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

† Departments of Pediatrics and Genetics, and the Howard Hughes Medical Institute, Beckman Center for Molecular and Genetic Medicine, Stanford University Medical Center, Stanford, California 94305-5428, USA

§ Department of Biochemistry, The University of Hong Kong, 5 Sassoon Road, Hong Kong

Hox genes control regional identity during segmentation of the vertebrate hindbrain into rhombomeres^{1–3}. Here we use transgenic analysis to investigate the upstream mechanisms for regulation of *Hoxb-3* in rhombomere(r)5. We identified enhancers from the mouse and chick genes sufficient for r5-restricted expression. Sequence comparisons revealed two blocks of similarity (of 19 and 45 base pairs), which each contain *in vitro* binding sites for the *kreisler* protein (Krm11), a Maf/b-Zip protein expressed in r5 and r6 (ref. 4). Both sites are required for r5 activity, suggesting that *Hoxb-3* is a direct target of *kreisler*. Multimers of the 19-base-pair (bp) block recreate a Krm11-like pattern in r5/r6, but the 45-bp block mediates expression only in r5. Therefore elements within the 45-bp block restrict the response to Krm11. We identified additional sequences that contain an Ets-related activation site, required for both the activation and restriction to r5. These studies demonstrate that *Krm11* directly activates expression of

† Present address: Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5 Canada.

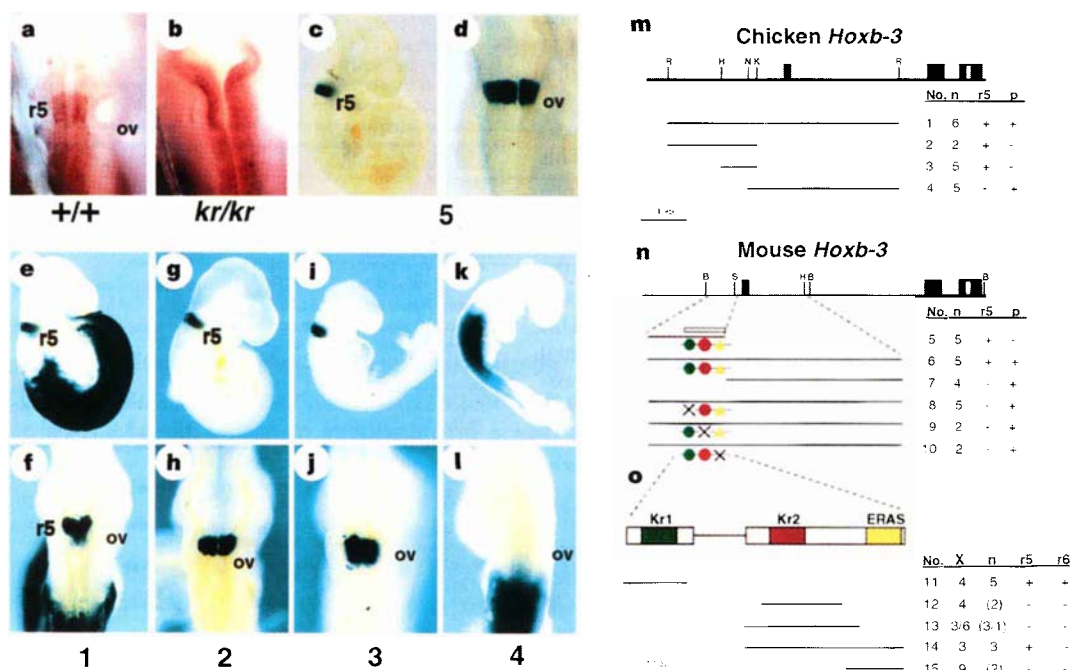


Figure 1 Transgenic analysis of *Hoxb-3* r5 regulatory regions. **a, b**, *Hoxb-3* expression in wild-type (**a**) and *kreisler* embryos (**b**). **c, d**, Lateral and dorsal views, respectively, showing mouse r5 enhancer activity (construct 5). **e-l**, Lateral (**e, g, i, k**) and dorsal (**f, h, j, l**) views of expression directed by chick r5 enhancer constructs (**m**). **m**, Chick *Hoxb-3* genomic region and *lacZ* reporter constructs used in deletion analysis. **n**, Mouse *Hoxb-3* locus and transgenic constructs. White bar above construct 5 represents the fragment used for DNase footprinting (Fig. 2c). **o**, Top, the conserved mouse sequence shown in Fig. 2a; below,

constructs with multimerized oligonucleotides (X, number of copies per construct). Numbers under panels represent construct number; n, total number of transgenic embryos with correct expression; numbers in brackets, number of embryos with ectopic expression; p, expression in posterior mesoderm; ov, otic vesicle; B, *Bam*HI; H, *Hind*III; K, *Kpn*I; N, *Not*I; R, *Eco*RI; S, *Stu*I. Colour code for **n, o** as in Fig. 3. All embryos are 9–9.5 d.p.c.; black boxes, exons; and white boxes, homeobox.

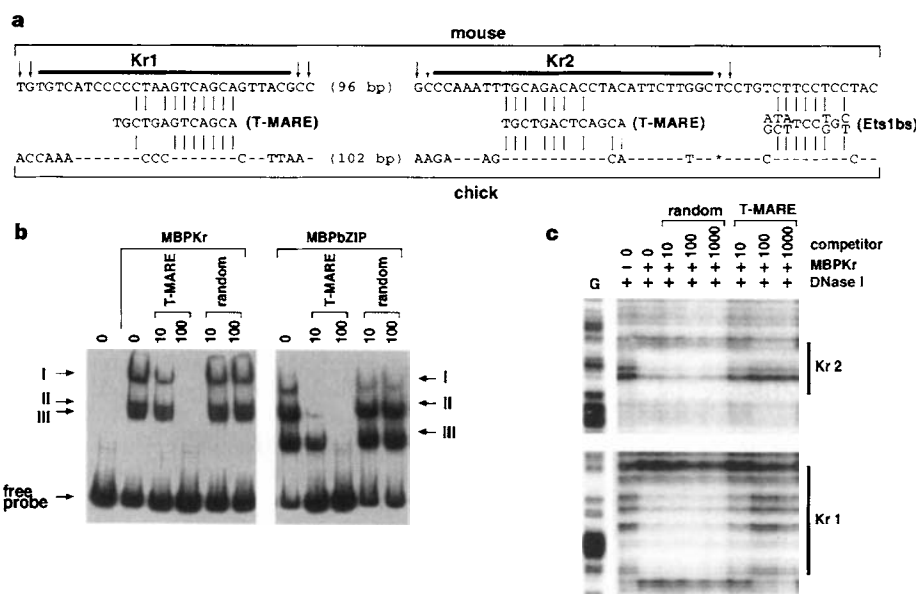


Figure 2 Binding of recombinant Krm1 protein to the mouse r5 enhancer. **a**, Alignment of sequences for the mouse and chick r5 enhancers, with identity to consensus binding sites for Maf (T-MARE) and Ets-1 (Ets1bs) proteins. Bars indicate the Kr1 and Kr2 MBPKr footprints; arrows, DNase I hypersensitive sites. **b**, Electrophoretic mobility shift of the 650-bp r5 enhancer by recombinant Krm1 proteins. Three protein/DNA complexes are detected (I, II, III) and binding of MBPKr or MBPzIP proteins is competed by a molar excess (10 ×, 100 ×)

of a T-MARE but not a random oligonucleotide. **c**, DNase I footprinting analysis of a 329-bp fragment from the mouse 650-bp r5 enhancer (Fig. 1n) alongside a Maxam-Gilbert sequencing reaction (G) as a size marker. Addition of DNase I, MBPKr protein or competitor oligonucleotides is indicated above (+ or – signs). Vertical bars on the right mark the protected sites and correspond to the reverse complement of the sequences in **a**.

***Hoxb-3* in r5 in combination with an Ets-related activation site, and suggest that *kreisler* plays a primary role in regulating segmental identity through *Hox* genes.**

At 9.0 days post coitum (d.p.c.), *Hoxb-3* is normally expressed in posterior mesoderm, in neuroepithelium caudal to the r5/r6 boundary at low levels, and at high levels within r5 (ref. 5) (Fig. 1a). However, in three independent mouse mutants, the radiation-induced *kreisler*⁶⁻⁹ and targeted *Krox-20* (refs 10, 11) and *Hoxa-1* (refs 12-14) genes, there are segmental abnormalities in the hindbrain and r5 is lost on the basis of changes in marker gene expression, such as the upregulation of *Hoxb-3* (Fig. 1b). These changes could be direct or indirect owing to the loss of this segment, and to distinguish between these possibilities we investigated the *cis* requirements for expression of *Hoxb-3* in r5. Genomic fragments from the chick and mouse *Hoxb-3* genes were linked to a *lacZ* reporter gene and tested in transgenic mice. Initially using deletion analysis of the chick locus (Fig. 1e-m), we identified an 800-bp fragment that mediates r5-restricted expression in a manner analogous to the endogenous mouse *Hoxb-3* gene (construct 3; Fig. 1i,j). In an equivalent 5' flanking region of the mouse gene, a 650-bp fragment with r5 regulatory activity was also identified (construct 5; Fig. 1c,d,n). Sequence comparison of the chick and mouse r5 enhancers revealed only two short segments of similarity, 19 and 45 bp, separated by ~100 bp. A 180-bp fragment containing these two conserved blocks and the spacer region was sufficient to direct r5-restricted expression (data not shown). No obvious consensus *Krox-20* binding sites of the type found in *Hoxa-2* and *Hoxb-2* (ref. 15) were present in the short conserved blocks, and ectopic *Krox-20* expression in transgenic embryos (*n* > 60) failed to transactivate the mouse *Hoxb-3* r5 enhancer (data not shown). However, the 19- and 45-bp blocks each contained partial 13-bp consensus binding sites for Maf proteins¹⁶ (T-MARE; Fig. 2a). As *kreisler* (*Krml1*) is a Maf family member⁴, we tested whether it could be directly involved in upregulation of *Hoxb-3* in r5 through these motifs.

To determine whether *Krml1* would bind to these and/or additional sites within the r5 enhancer, we did electrophoretic mobility shift assays and DNase I footprinting experiments using recombinant maltose-binding protein fused to sequences from *Krml1* (MBPKr). Incubation with the mouse 650-bp r5 enhancer produced three distinct protein/DNA complexes with reduced electrophoretic mobility compared to DNA alone (Fig. 1n, 2b). A fusion protein that contained only the basic domain/leucine-zipper region of *Krml1* (MBPbZIP) produced a similar gel-shift pattern, and binding was blocked specifically by an excess of T-MARE double-stranded oligonucleotide (Fig. 2b), which suggested that binding of *Krml1* to DNA occurred in a manner similar to that characterized previously for other Maf-related proteins^{16,17}. In DNase I footprinting experiments, recombinant MBPKr protein protected 27- and 29-bp regions, termed the Kr1 and Kr2 sites, respectively (Fig. 2a,c). Each site contained a central region resembling the T-MARE consensus, and the footprints were blocked specifically by an excess of T-MARE oligonucleotide (Fig. 2a,c), or by an oligonucleotide from the predicted chick Kr1 site (data not shown). The 19-bp segment of similarity between mouse and chick r5 enhancers was contained completely within Kr1; however, the 45-bp segment of similarity extended beyond Kr2.

The role of Kr1 and Kr2 in r5 expression *in vivo* was determined by constructing 7-bp deletions in each site that removed the most conserved consensus T-MARE half-site (5'-GTCAGCA for Kr1; 5'-TGCAGAC for Kr2; Fig. 2a), and testing the variants in transgenic embryos (Fig. 1n). As an internal control, the deletions were generated in the context of a larger 2.2-kilobase (kb) *Bam*HI-*Hind*III fragment from the mouse gene that mediates expression in posterior mesoderm in addition to r5 (construct 6; Fig. 3a,f, and C.-T.K. and M.-H.S., unpublished results). Deletion of either Kr1 or Kr2 specifically abolishes expression only in r5 (constructs 8 and 9; Fig. 3c,d,h,i), in a way identical to that seen

when the entire 650-bp r5 enhancer is deleted (construct 7; Fig. 3b,g). Additional transgenic studies encompassing >45 kb surrounding the *Hoxb-3* gene have not identified any elements other than this 650-bp fragment capable of mediating r5-restricted expression (data not shown). Taken together, these results indicate that Kr sites 1 and 2 are essential components of the regulatory mechanism governing high-level expression of *Hoxb-3* in r5.

To ascertain whether Kr1 or Kr2 sites are sufficient to mediate r5-restricted expression, tandem copies of double-stranded oligonucleotides covering each site were tested for their ability to direct expression of *lacZ* in transgenic mice (Fig. 1o). Four copies of the Kr1 oligonucleotide generated reporter expression not only in r5, but also in the rostral half of r6 and its associated neural crest (construct 11; Fig. 4b,f). This expanded pattern is identical to that of *Krml1* RNA (Fig. 4a,e), which suggests that the multimerized Kr1 site provides a direct readout of endogenous *Krml1*. As expression directed by the multimerized Kr1 site differs from that of the 650-bp *Hoxb-3* enhancer, in that it is not restricted only to r5, our results imply that other components are normally limiting the *Hoxb-3* response to *Krml1* in r6.

In contrast to Kr1, multimers of two different oligonucleotides spanning the Kr2 site failed to produce any rhombomere-restricted expression (constructs 12 and 13; Fig. 4c,g). However, when the oligonucleotide-spanning Kr2 site was lengthened by 12 bp to include the entire 45-bp segment conserved between the mouse and chick r5 enhancers, a three-copy multimer produced *lacZ* expression restricted to r5 (construct 14; Fig. 4d,h). This pattern is not the simple readout of *Krml1* expression as seen with Kr1 site (Fig. 4b,f), but is identical to that observed with the entire mouse and chick r5 enhancers (Fig. 1c,d,i,j). To investigate whether the 12-bp region immediately adjacent to the Kr2 site was itself responsible for r5 expression, we tested a nine-copy multimer that spanned this region and found no *lacZ* expression (construct 15; data not shown). These results suggest that the conserved 45-bp segment contains sequences within the 12 bp at its 3' end that serve to both potentiate and restrict the ability of *Krml1* to direct r5 expression *in vivo* from Kr2.

Within this 12-bp 3' region there is a 9-bp partial consensus binding site for the Ets-1 transcription factor (Fig. 2a)¹⁸. Gel-shift

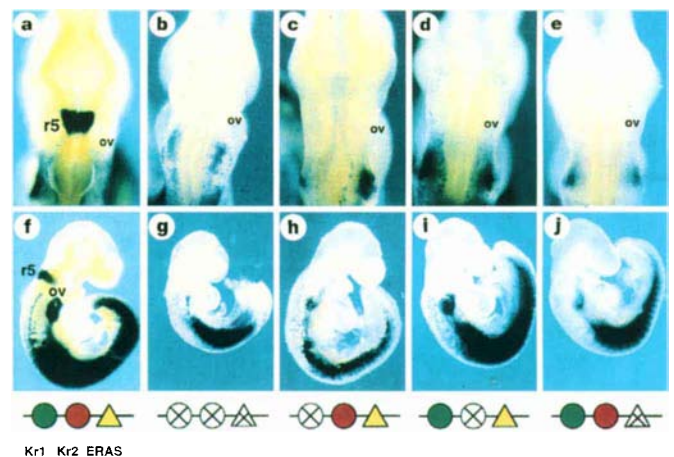


Figure 3 The *Krml1* and the Ets related activation sites (ERAS) are required for r5 enhancer activity. **a, f**, Transgenic embryos for the 2.2-kb *Bam*HI-*Hind*III mouse enhancer (construct 6) showing staining in r5 and posterior mesoderm. **b, g**, Staining in embryo with transgene carrying deletion of all three sites (construct 7). **c-e** and **h-j**, Expression mediated by variants with mutations in either the Kr1 site (construct 8; **c, h**), the Kr2 site (construct 9; **d, i**) or the ERAS (construct 10; **e, j**). In all three cases, r5 expression is specifically lost. Green circle, Kr1 site; red circle, Kr2 site; yellow triangle, ERAS; white and crossed elements indicate deletion or mutation of the site (Fig. 1n). All embryos are 9-9.5 d.p.c.; **a-e**, dorsal; **f-j**, lateral views; ov, otic vesicle.

assays confirmed that the r5 enhancer and an oligonucleotide containing this predicted Ets-related activation site (ERAS) bind specifically to recombinant Ets-1 protein (data not shown). We investigated whether or not r5 enhancer function would be affected by point mutations that changed the ERAS 5'-GGAGGA core to 5'-AAGCTT, which are alterations that will abolish Ets domain binding activity *in vitro*^{18,19}. In the context of the 2.2-kb *Hoxb-3* enhancer, this mutation eliminated expression of *lacZ* in r5, in a way identical to that observed when the Kr1 or Kr2 sites are deleted, but had no effect in posterior mesoderm (construct 10; Fig. 3e,j). As the ERAS is required for full enhancer activity, it is not simply a repressor involved in blocking expression in r6 because transgene staining is also lost in r5. Therefore, our results indicate that a transcription factor(s) with binding activity similar to Ets-1 or overlapping with the ERAS is essential for activation of *Hoxb-3* in r5, and suggest that this factor may cooperatively interact with Krml1 to restrict its potential for transactivation *in vivo*. The idea that this factor might be a member of the large vertebrate Ets-domain protein family (>20 members) is supported by studies demonstrating cooperativity between AP-1 like (b-Zip) and Ets-like factors^{20,21}. In addition, Ets-1 is known to interact with MafB (the chicken Krml1 homologue) in regulating lineage decisions during avian haematopoiesis²². However, Ets-1 itself cannot be the member required for r5 activation through the ERAS because it is not expressed in r5, but it is detected in even-numbered rhombomeres (data not shown)²³.

Our results demonstrate that multiple Krml1 sites in the enhancer are required for r5 activity, as mutation of either Kr1 or Kr2 abolishes expression. Despite their similar *in vitro* binding properties, the Kr1 and Kr2 sites are not equivalent *in vivo*. Multimers of Kr1 mediate a full readout of endogenous *kreisler*, whereas Kr2 does not. We have uncovered a requirement for additional factors that potentiate Kr2 and restrict its response to r5. This suggests a model in which Krml1 binds to the Kr2 site with low efficiency, but is recruited to the DNA by factors interacting with the adjacent ERAS. Spatial restriction of these cofactors or their activities could also serve to restrict the ability of Krml1 to activate *Hoxb-3* expression in r5 compared with r6. Alternatively, the restriction to r5 could involve the other conserved sequences in the 45-bp block. Together, these results show that *Hoxb-3* is a direct target of *kreisler* in the

cascade controlling patterning of the vertebrate hindbrain, and suggest that *kreisler* plays a primary role in regulating segmental identity through *Hox* genes. This indicates that developmental abnormalities in *kr/kr* embryos are likely to be caused by a failure of cells in the caudal hindbrain to acquire their normal r5 identity owing to abnormal domains of *Hox* expression. □

Methods

Mice and constructs. Transgenic mice were produced and assayed for *lacZ* activity as described²⁴. Constructs were generated in a *lacZ* vector with the human β -globin promoter (pBGZ40²⁵; constructs 1–4, 11–15) or the mouse *Hoxb-4* promoter²⁴ (constructs 5–10). Deletions and mutations were generated by site-directed mutagenesis in m13 and were checked by sequencing. Double-stranded oligonucleotides with overhanging *SpeI*-compatible ends were cloned in the *SpeI* site of pBGZ40 and the copy number determined by sequencing. Sequence of the oligonucleotides used was: construct 11, 5'-TCCCCCTAAGT-CAGCAGTT; construct 12, 5'-TGCAGACACCTACATTCTTGGC; construct 13, 5'-AAATTTGCAGACACCTACATTCTTGGCTCTG; construct 14, 5'-AAATTTGCAGACACCTACATTCTTGGCTCTGCTCTCTCTCTACAG; construct 15, 5'-CCTGTCTCTCTCTCTACAG. Probes used for *in situ* hybridization and maintenance and genotyping of *kr/kr* mice have been described^{14,26}.

Electrophoretic mobility shift assays. The MBPKr and KBpZIP fusion proteins contain amino-acid residues 18–323 and 208–323, respectively, of the mouse *kreisler* sequences⁴, fused to sequences of the pMAL-c vector (New England Biolabs). Fusion proteins were purified at 99% as estimated by SDS-PAGE and resuspended at 700 ng ml⁻¹. DNA-binding reactions contained 250 pg DNA radiolabelled with the Klenow fragment of DNA polymerase and either 700 ng MBPKr protein or 100 ng MBpZIP protein in a final volume of 15 μ l containing 20 mM HEPES, pH 7.9, 1 mM EDTA, 20 mM KCl, 5 mM DTT, 4 mM MgCl₂ and 200 μ g ml⁻¹ poly d(I-C). Competitor double-stranded T-MARE (5'-AGCTCGGAATTGTGTGACTGACGATTACTC) or random (5'-GAGTAATGAGGACTCTCAATTCGAG-3') oligonucleotide was added at the start of the reaction. After 30 min incubation at room temperature, reactions were subjected to electrophoresis on 4% non-denaturing polyacrylamide gels and analysed by autoradiography.

DNase I footprinting assays. A subclone of 329 bp was linearized with *HindIII* at the left-hand end of the vector/insert junction, end-labelled with the Klenow fragment of DNA polymerase, and the released by digestion with *SacII*. DNA-binding reactions were carried out as for the gel shift assay, and where indicated, 200 ng DNase I was added for 1 min, and the reactions were electrophoresed on 6% denaturing polyacrylamide gels.

Received 2 December 1996; accepted 14 March 1997.

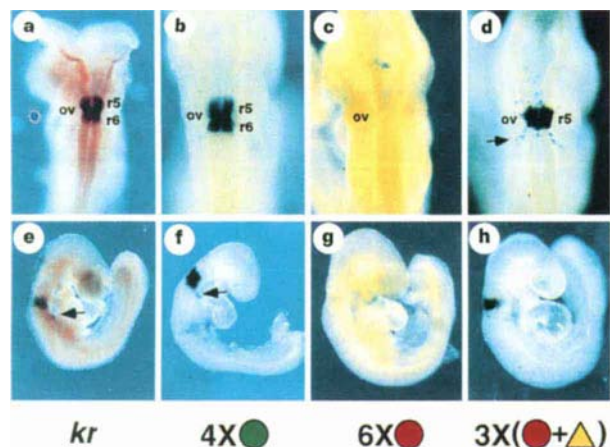


Figure 4 Differential expression mediated by the conserved blocks of the r5 enhancer. **a, e, h**, *kreisler* RNA expression in r5, rostral r6 and neural crest from r5/6 (arrow). **b, f**, Four copies of the Kr1 site (construct 11) give reporter staining similar to endogenous *kreisler*. **c, g**, Multimers of Kr2 site (construct 13) do not restrict expression in the neural tube. **d, h**, Three copies of the 45-bp conserved block (construct 14) give r5-restricted expression. Green circle, Kr1 site; red circle, Kr2 site; yellow triangle, ERAS. All embryos are 9–9.5 d.p.c. except for **a** (8.75–9 d.p.c.). ov, otic vesicle. Arrows indicate migrating neural crest.

- Lumsden, A. & Krumlauf, R. Patterning the vertebrate neuraxis. *Science* **274**, 1109–1115 (1996).
- Studer, M., Lumsden, A., Ariza-McNaughton, L., Bradley, A. & Krumlauf, R. Altered segmental identity and abnormal migration of motor neurons in mice lacking *Hoxb-1*. *Nature* **384**, 630–634 (1996).
- Goddard, J., Rossel, M., Manley, N. & Capecchi, M. Mice with targeted disruption of *Hoxb1* fail to form the motor nucleus of the VIIth nerve. *Development* **122**, 3217–3228 (1996).
- Cordes, S. P. & Barsh, G. S. The mouse segmentation gene *kr* encodes a novel basic domain-leucine zipper transcription factor. *Cell* **79**, 1025–1034 (1994).
- Hunt, P. et al. A distinct *Hox* code for the branchial region of the head. *Nature* **253**, 861–864 (1991).
- Deol, M. The abnormalities of the inner ear in *kreisler* mice. *J. Embryol. Exp. Morphol.* **12**, 475–490 (1964).
- Frohman, M. A., Martin, G. R., Cordes, S., Halamek, L. P. & Barsh, G. S. Altered rhombomere-specific gene expression and hyoid bone differentiation in the mouse segmentation mutant *kreisler* (*kr*). *Development* **117**, 925–936 (1993).
- McKay, I. J. et al. The *kreisler* mouse: a hindbrain segmentation mutant that lacks two rhombomeres. *Development* **120**, 2199–2211 (1994).
- McKay, I. & Lumsden, A. Motor neuron defects in the *kreisler* mutant mouse. *Eur. J. Neurosci.* (in the press).
- Swiatek, P. J. & Gridley, T. Perinatal lethality and defects in hindbrain development in mice homozygous for a targeted mutation of the zinc finger gene *Krox-20*. *Genes Dev.* **7**, 2071–2084 (1993).
- Schneider-Maunoury, S. et al. Disruption of *Krox-20* results in alteration of rhombomeres 3 and 5 in the developing hindbrain. *Cell* **75**, 1199–1214 (1993).
- Carpenter, E. M., Goddard, J. M., Chisaka, O., Manley, N. R. & Capecchi, M. R. Loss of *Hoxa-1* (*Hox-1.6*) function results in the reorganization of the murine hindbrain. *Development* **118**, 1063–1075 (1993).
- Dolle, P. et al. Local alterations of *Krox-20* and *Hox* gene expression in the hindbrain of *Hox-1* (*Hox-1.6*) homozygote null mutant embryos. *Proc. Natl Acad. Sci. USA* **90**, 7666–7670 (1993).
- Mark, M. et al. Two rhombomeres are altered in *Hoxa-1* mutant mice. *Development* **119**, 319–338 (1993).
- Nonchev, S. et al. The conserved role of *Krox-20* in directing *Hox* gene expression during vertebrate hindbrain segmentation. *Proc. Natl Acad. Sci. USA* **93**, 9339–9345 (1996).
- Kataoka, K., Noda, N. & Nishizawa, M. Maf nuclear oncoprotein recognizes sequences related to an AP-1 sites and forms heterodimers with both Fos and Jun. *Mol. Cell. Biol.* **14**, 700–712 (1994).
- Kataoka, K., Fujiwara, K. T., Noda, M. & Nishizawa, M. MafB, a new Maf family transcription activator that can associate with Maf and Fos but not with Jun. *Mol. Cell. Biol.* **14**, 7581–7591 (1994).

18. Nye, J., Petersen, J., Gunther, C., Jonsen, M. & Graves, B. Interaction of murine Ets-1 with GGA-binding sites establishes the ETS domain as a new DNA binding motif. *Genes Dev.* **6**, 975–990 (1992).
19. Shore, P. et al. Determinants of DNA binding specificity of ETS-domain transcription factors. *Mol. Cell. Biol.* **16**, 3338–3349 (1996).
20. Wasyluk, B. et al. The *c-ets* proto-oncogenes encode transcription factors that cooperate with c-Fos and c-Jun for transcriptional activation. *Nature* **346**, 191–193 (1990).
21. Treier, M., Bohmann, D. & Mlozik, M. JUN cooperates with the Ets domain protein Pointed to induce photoreceptor R7 fate in the *Drosophila* eye. *Cell* **83**, 753–776 (1995).
22. Sieweke, M., Tekotte, H., Frampton, J. & Graf, T. MafB is an interaction partner and repressor of Ets-1 that inhibits erythroid differentiation. *Cell* **85**, 49–60 (1996).
23. Maroulakou, I., Papas, T. & Green, J. Differential expression of *ets-1* and *ets-2* proto-oncogenes during murine embryogenesis. *Oncogene* **9**, 1551–1565 (1994).
24. Whiting, J. et al. Multiple spatially specific enhancers are required to reconstruct the pattern of *Hox-2.6* gene expression. *Genes Dev.* **5**, 2048–2059 (1991).
25. Yee, S.-P. & Rigby, P. W. J. The regulation of *myogenin* gene expression during the embryonic development of the mouse. *Genes Dev.* **7**, 1277–1289 (1993).
26. Sham, M.-H. et al. Analysis of the murine *Hox-2.7* gene: conserved alternative transcripts with differential distributions in the nervous system and the potential for shared regulatory regions. *EMBO J.* **11**, 1825–1836 (1992).

Acknowledgements. We thank K. Maruthainar for sequencing; A. Kuroiwa for chicken clones; L. Ariza-McNaughton for help with *in situ* hybridization; B. Graves for the gift of Ets-1 protein; Z. Webster and A. Hewett for animal husbandry; and all other members of R.K.'s laboratory for advice and encouragement. This work was supported by fellowships from EMBO and HFSP (M.M.) by the MRC and an HFSP network grant (R.K.); by grants from the Hong Kong RGC (C.-T.K. and M.H.S.); and by a fellowship from the ACS (S.C.) and a grant from the NIH (S.B.); G.S.B. is an assistant investigator of the HHMI.

Correspondence and requests for materials should be addressed to R.K. (e-mail: r-krumlauf@nimr.mrc.ac.uk).

Sex lethal controls dosage compensation in *Drosophila* by a non-splicing mechanism

Richard L. Kelley*, Jiwu Wang†, Leslie Bell† & Mitzi I. Kuroda*‡

* Department of Cell Biology and the ‡ Howard Hughes Medical Institute, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA

† Molecular Biology Program, University of Southern California, 835 West 37th Street, Los Angeles, California 90089-1340, USA

Dosage compensation in *Drosophila* requires the male-specific lethal (*msl*) proteins (MSL) to make gene expression from the single male X chromosome equivalent to that from both female X chromosomes^{1,2}. Expression of *msl2* is repressed post-transcriptionally by Sex lethal (SXL), a female-specific RNA-binding protein that regulates alternative splicing in the sex-determination hierarchy. Although *msl2* RNA is alternatively spliced in males and females, this does not alter its coding potential and splicing is not required for male-specific expression of MSL2 protein. Instead, our results suggest that the association of SXL protein with multiple sites in the 5' and 3' untranslated regions of the *msl2* transcript represses its translation in females. Thus, this well characterized alternative splicing factor regulates at least one target transcript by a distinct mechanism.

Male-specific lethal-2 (*msl2*) encodes a key male-limited regulator of dosage compensation in *Drosophila*. The MSL proteins function as a multisubunit complex bound to hundreds of sites on the male X chromosome, where they are postulated to recruit the MOF (males absent on the first) histone acetyltransferase³. Histone acetylation is thought to play a major role in remodelling the chromatin architecture of the male X chromosome, allowing a doubling in transcription². The MSL complex is essential for male development and is normally absent from females. Inappropriate expression of high levels of MSL complex is toxic to females because of hyperexpression of both X chromosomes⁴. This critical binary switch is regulated by the *Sex lethal* (*Sxl*) gene^{5,6} and its primary target in this dosage compensation pathway is the *msl2* transcript^{4,7,8}. SXL is a female-specific RNA-binding protein that regulates the sex-specific splicing of its own transcript as well as that of *transformer* in the sex determination hierarchy^{9,10}. Therefore, it was reasonable to presume that SXL might regulate *msl2* using a similar mechanism.

The *msl2* gene produces two transcripts that differ by an intron of 133 nucleotides in the 5' untranslated region (UTR) (Fig. 1)^{4,7,8}. Most female transcripts retain the intron, whereas most male transcripts remove it (Fig. 1b)⁷. Consensus binding sites for SXL protein (AU₇ or U₈₊)¹¹ are located adjacent to both 5' and 3' splice junctions (Fig. 1c)^{4,7,8}, and female cells that are mutant for *Sxl* splice the 5' intron⁸ and derepress *msl2* translation⁴. However, unlike the previously described cases of SXL regulation, this sex-specific intron is located 279 nucleotides before the AUG start codon and does not affect the open reading frame. In addition to the size difference, *msl2* RNA is less abundant in females than males⁷, but this difference is not sufficient to account for the absence of MSL2 protein in females. Instead, it has been proposed that *msl2* is repressed at the level of translation in females^{4,7,8}. We have explored two models for sex-specific regulation of *msl2* expression by *Sxl*. In the first model, some feature of the *msl2* intron, such as RNA secondary structure or a non-productive initiation codon, results in translational inhibition. If so, male viability would require removal of the intron so that MSL2 translation could occur. SXL would play an indirect role by blocking removal of the intron from female transcripts. An alternative model is that translational inhibition results from direct association of SXL protein with multiple poly(U) sites within the 5' and 3' UTRs of the mature transcript in females. According to this model, there is no obvious reason to splice out the intron in males because they lack the SXL repressor protein.

To distinguish between these models, we constructed a series of wild-type and mutant *msl2* transgenes (Fig. 1) and introduced them into flies lacking a functional *msl2* gene. We used western blots and the more sensitive fluorescent immunolocalization assay on polytene chromosomes to measure MSL2 levels *in vivo*. All constructs carry an 11-nucleotide *Bgl*II linker inserted 22 nucleotides before the AUG start codon. This eliminates the possibility that any MSL2 protein might arise from translation initiation upstream in the long 5' UTR and also allows endogenous and transgenic *msl2* RNAs to be distinguished. The wild-type control transgene, M2NG11, is properly regulated such that MSL2 protein is made in males but not females (Fig. 2b, d), and its transcript is properly spliced in each sex (Fig. 3a,b, lanes 3, 4). The second transgene, SM53, carries mutations in both 5' and 3' splice junctions of the sex-specific intron (Fig. 1b, c). Reverse-transcribed polymerase chain reaction (RT-PCR) analysis of *msl2* RNA derived from this construct shows that both males and females produce identical transcripts retaining the intron in the 5' UTR (Fig. 3b, lanes 5, 6). This 'female-like' *msl2* RNA is efficiently translated in males, as shown by (1) rescue of *msl2* males, (2) wild-type pattern of MSL2 protein distribution on the transgenic male X (Fig. 2f), and (3) similar levels of MSL2 protein in SM53 and M2NG11 transgenic males when assayed by western blots (Fig. 2s). These results demonstrate that the intron within the 5' UTR does not contain sequences that inherently block translation. Besides the intron, the SM53 transcript carries a long 5' UTR with several short opening reading frames starting with AUG. Such structures can strongly inhibit cap-dependent translation initiation in many organisms^{12,13}, but are commonly encountered in *Drosophila*¹⁴. The ability of males to translate the SM53 *msl2* RNA shows that this 5' structure is insufficient to block translation at the authentic AUG in *Drosophila*. By contrast, the same SM53 transcript produces no protein in females that contain the SXL repressor (Fig. 2h). Thus, SXL is able to regulate SM53 *msl2* translation properly in the absence of alternative splicing. We conclude that SXL does not require the presence of splice junctions to be able to recognize its target.

The next *msl2* transgene tested, SXB1, contains the same splice junction mutations as SM53, with additional changes in the first poly(U) segment postulated to bind SXL protein, located just inside the male intron (Fig. 1b, c). This mutant transgene functions properly in males. Western analysis failed to detect any MSL2 protein in females, but examination of polytene chromosomes

revealed that translation is weakly derepressed (Fig. 2j). Mutations in the second poly(U) segment (SXB2) also allow modest levels of translation in females (Fig. 2l). This demonstrates that complete repression in females requires a full set of SXL binding sites. Transgenes mutant for both poly(U) segments in the 5' UTR (SXB1-2) show a greater derepression in females (Fig. 2n), which is high enough to be detected by western blots (Fig. 2s).

The *msl2* transcript contains four poly(U) stretches in the 3' UTR in addition to the two sites found within the male-specific intron in the 5' UTR (Fig. 1). Removal of the 3' UTR was previously shown to cause weak derepression in females⁸. We replaced the 3' UTR with the polyadenylation region of simian virus 40 (SV40) in pMM2SV, and also observed weak derepression in females (Fig. 2p). When all of the poly (U₇) stretches are mutated or deleted in the NOPU transgene, high MSL2 levels are observed in females, indicating similar translational efficiency in males and females (Fig. 2r). Females carrying one copy of NOPU contain about 30% of the MSL2 protein found in their transgenic brothers (Fig. 2s). This amount of MSL2 is insufficient to kill females, possibly because they have twice as much X chromatin to act upon and also because MSL1 is downregulated in females^{15,16}. We therefore constructed females carrying two copies of the NOPU transgene and one copy of a transgene overexpressing MSL1 (K. Chang and M.I.K., unpublished). Such females have delayed development because of inappropriate dosage compensation (data not shown).

We next examined whether the reduced levels of *msl2* RNA in females are a consequence of SXL action. Quantitative nuclease protection assays revealed that wild-type females contain only about 20% as much *msl2* RNA as males (Fig. 3c,d). However, female *msl2*

transcripts lacking SXL-binding sites made by the NOPU transgene accumulate to about 70% of male levels when normalized to *rp49* transcripts (Fig. 3d). This suggests that translated *msl2* RNA is more stable than transcripts repressed by SXL, although we have not excluded the possibility that mutating the poly(U) clusters somehow stimulates transcription preferentially in females.

To investigate whether SXL repression of *msl2* could be direct, we tested SXL protein for binding to *msl2* transcripts *in vitro*. Figure 4a shows that two shifted bands are evident when SXL protein is incubated with a 246-nucleotide region of the *msl2* 5' UTR containing both poly(U) stretches. Similarly, two shifted bands are also observed with a poly(U)-containing region of the *msl2* 3' UTR. Overall binding is stronger to the 5' UTR than the 3' UTR, on the basis of competition experiments with cold Sxl RNA containing a U₉AU₈ sequence (data not shown). Figure 4b compares the SXL binding ability of a 340-nucleotide region of the *msl2* 5' UTR (M2NG11) with the mutations used in the *in vivo* analysis. The unaltered RNA shows two shifted bands, and mutations in the splice junctions of the male intron (SM53) do not affect SXL binding. However, consistent with the derepression seen *in vivo*, mutating individual poly(U) segments reduces SXL binding *in vitro*. Altering the U₁₁ run (SXB1) causes a modest reduction, altering the U₁₆ run (SXB2) causes a significant reduction, and transcripts with mutations in both poly(U) runs (SXB1-2) fail to bind SXL at the concentrations tested.

Translational control is emerging as a widely used regulatory mechanism in higher organisms. The expression of several *Drosophila* and *Caenorhabditis elegans* proteins required for embryonic pattern formation, sex determination, and germ-cell

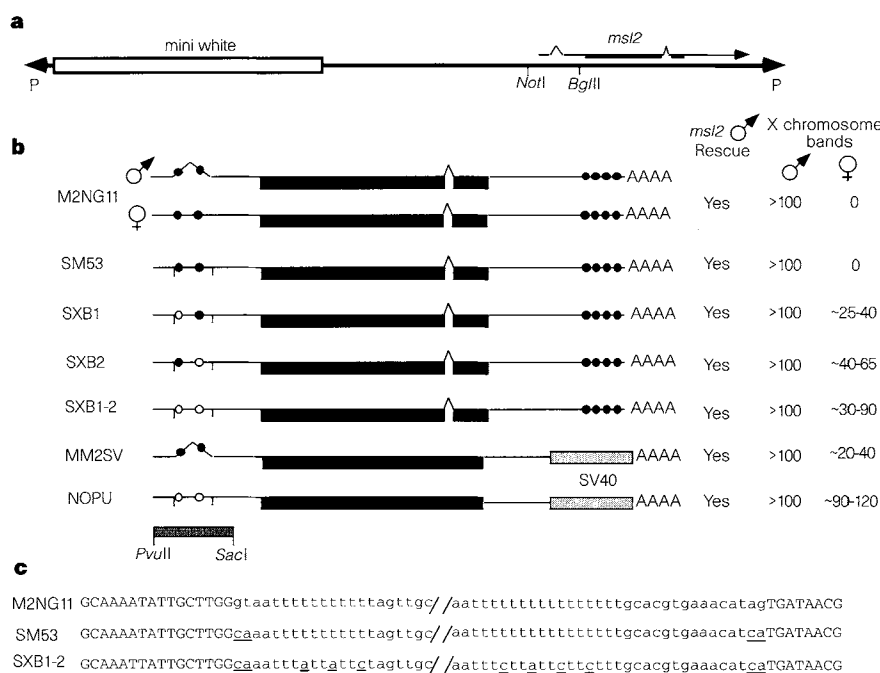


Figure 1 Summary of *in vivo* analysis of *msl2* regulation. **a**, All transgenes carried a complete *msl2* gene (6 kb) cloned into pCasPer (ref. 21). *Clal* and *XmaI* sites at the 5' end of *msl2* were converted with oligomer adapters to *NotI* and *BglII* sites, respectively, to provide unique cloning sites. **b**, M2NG11 carries the wild-type *msl2* gene. Females retain the 5' intron, whereas males splice it out of most transcripts. The coding region is shown as a thick bar and the 5' and 3' UTRs as thin lines. Consensus SXL binding sites are black circles. Mutated sites are white circles. SM53, SXB1, SXB2, SXB1-2 and NOPU all carry mutations in both splice junctions of the male-specific intron shown as short vertical bars. SXB1 also has three mutations in the first poly(U) stretch; SXB2 has four nucleotide changes in the second poly(U) stretch; SXB1-2 and NOPU have both 5' poly(U) stretches mutated. MM2SV has a wild-type 5' UTR, but the poly(U) clusters in the 3' UTR

have been replaced by a segment of the SV40 genome containing a polyadenylation signal. NOPU carries the 5' end of SXB1-2 and the 3' end of MM2SV such that it lacks all poly(U) clusters. Each transgene was assayed for complementation of *msl2* mutant male lethality. In each case, the desired class of males was recovered in numbers equal to their *msl2* sisters, demonstrating complete rescue with several independent lines. The presence of multiple MSL2 bands on the female X chromosomes indicates a failure to repress *msl2* translation (Fig. 2). The shaded box at the bottom is the 340 bp *PvuII*-*SacI* fragment used to produce *in vitro* *msl2* transcripts for SXL gel shift experiments (Fig. 4). **c**, The sequences of the relevant portions of the *msl2* 5' UTR are shown with exon sequence in upper case and male intron sequence in lower case. The mutated nucleotides are underlined.

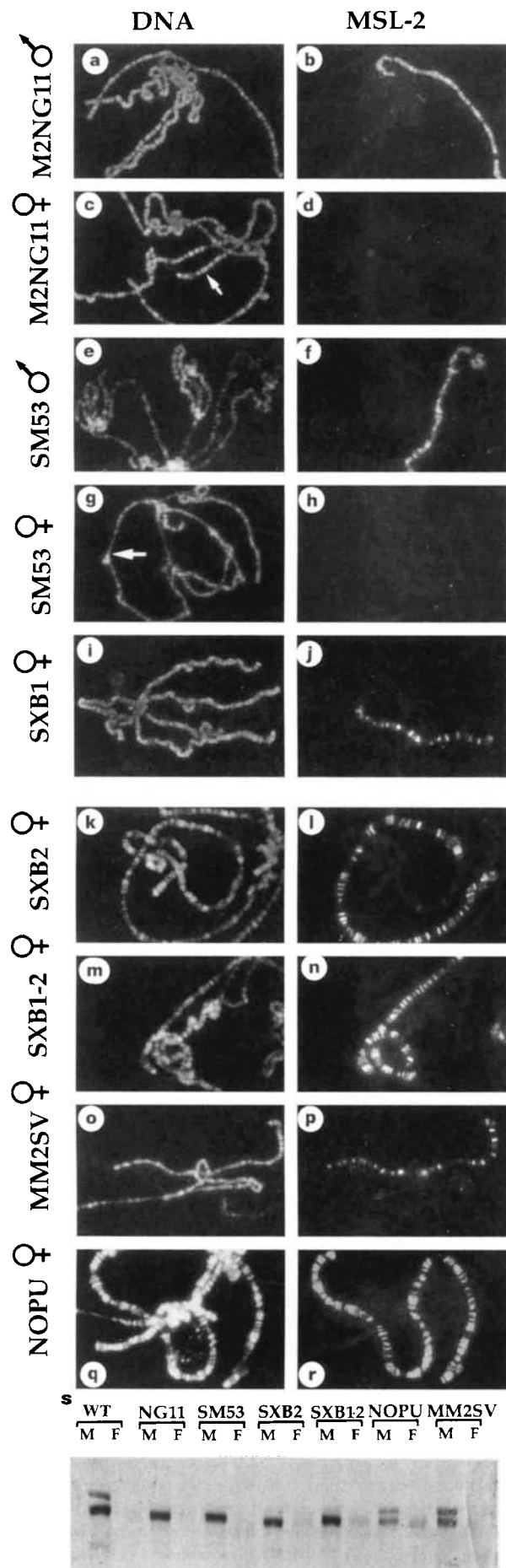


Figure 2 MSL2 expression in transgenics. Salivary gland chromosomes were stained for DNA (left) and MSL2 (right). All larvae were *w⁺ msl2^{cn}*, so the MSL2 protein detected must be derived from the single copy of the [*w⁺ msl2⁺*] transgene on either the X or 3rd chromosome. **a, b**, M2NG11 wild-type control male; **c, d**, M2NG11 wild-type control female; **e, f**, SM53 splice-mutant male; **g, h**, SM53 splice-mutant female; **i, j**, SXB1 poly(U) mutant female; **k, l**, SXB2 poly(U) mutant female; **m, n**, SXB1-2 poly(U) double-mutant female; **o, p**, MM2SV 3' UTR replacement female (carried two copies of transgene); **q, r**, NOPU females. The arrows in **c** and **g** indicate the female X chromosomes that fail to stain with MSL2 antibodies. **s**, Western blot to MSL2 protein in males (M) and females (F) of the same transgenic lines. The upper protein band seen in some lanes appears sporadically on western blots and is of unknown significance.

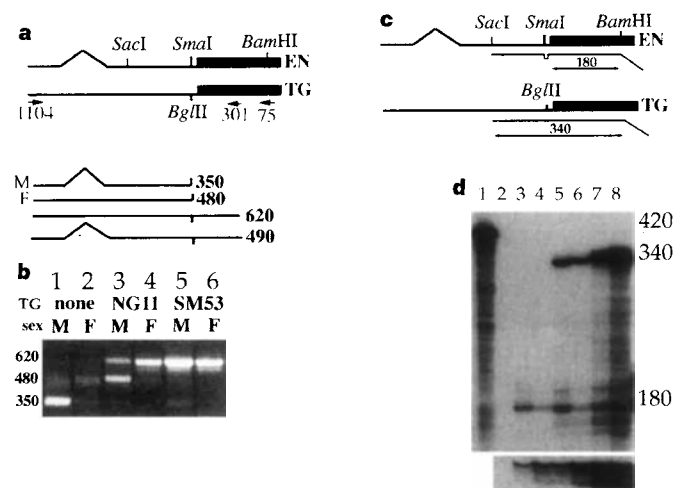


Figure 3 Analysis of transgenic *msl2* RNA. **a**, The 5' structures of endogenous (EN) and transgenic (TG) *msl2* RNA are shown along with the strategy for RT-PCR amplification. Thin lines are 5' UTR and thick lines are coding sequences. A *Sma*I site preceding the *msl2* start codon has been replaced with a *Bgl*II adapter in transgenes. Oligo number 75 was used for reverse transcription and PCR was carried out with oligos 1104 and 301. Following *Sma*I digestion, the endogenous male and female PCR products are 350 and 480 bp long, respectively. The transgenic products are not cut by *Sma*I, so spliced and unspliced molecules are 490 and 620 bp long, respectively. **b**, The endogenous *msl2* RNA from non-transgenic wild-type males (lane 1) and females (lane 2) produce predominantly spliced (350 bp) or unspliced (480 bp) PCR products, respectively. Animals carrying the control M2NG11 transgene with wild-type splice junctions show male-specific splicing of the transgenic transcript. The spliced M2NG11 male transgene product is 490 bp (lane 3) and the unspliced female transgene product is 620 bp (lane 4). Animals carrying the SM53 *msl2* splice mutant transgene produce the same unspliced product (620 bp) both in males (lane 5) and in females (lane 6). The 350-bp band seen in lane 5 is the product of the endogenous spliced male transcript. The endogenous transcript in the *msl2* host fly stock (lanes 3-6) is reverse-transcribed less efficiently than the transgenic product because primer 75 spans the proline 51 codon mutated in the *msl2¹* allele⁷. **c**, Nuclease protection assay of *msl2* RNA levels in wild-type and transgenic animals. The labelled riboprobe is shown below the transcript. The first 80 nt of the riboprobe (bent line) are derived from the pBlueScript vector and do not form base pairs with *msl2* RNA. The remainder is fully collinear with the transgenic sequence, but the *Bgl*II sequence does not form base pairs with endogenous *msl2* transcript. The endogenous *msl2* transcript (EN) protects a 180-nucleotide (nt) segment of the probe from nuclease digestion and a 340-nt fragment is protected by the transgenic RNA (TG). **d**, Results of the nuclease protection assay. Lane 1, starting probe; lane 2, nuclease digestion with no fly RNA; lane 3, with 1 μ g wild-type male poly(A)⁺ RNA; lane 4, with 1 μ g wild-type female poly(A)⁺ RNA; lane 5, with 1 μ g NOPU male poly(A)⁺ RNA; lane 6, with 1 μ g NOPU female poly(A)⁺ RNA; lane 7, with 3 μ g NOPU female poly(A)⁺ RNA; lane 8, with 10 μ g NOPU female poly(A)⁺ RNA. The lower panel shows the same RNA samples protecting a *rp49* probe, except that only 5% as much fly RNA was used in each assay and the exposure time was 7% as long as the upper panel. Endogenous female *msl2* RNA is about 500-fold less abundant than *rp49*.

formation are controlled at the level of translation by repressors binding to specific sequences in the 3' UTRs of their target transcripts¹⁷. The iron-dependent translational repressor binds to a stem-loop structure in the 5' UTR of its target transcripts¹⁸. We have shown that complete repression of *msl2* requires that SXL bind both the 5' and 3' UTRs. This may be important in light of a recent report that the 5' and 3' ends of transcripts are in close contact as messenger RNAs are loaded onto ribosomes¹⁹. The exact mechanism by which SXL represses translation of *msl2* RNA is not clear, as we have been unsuccessful in visualizing the spatial distribution of the low-abundance *msl2* RNA to see whether females sequester the transcript in the nucleus. Our results clearly demonstrate that the 5' UTR strongly inhibits translation, but only in the presence of SXL repressor. Although this provides a compelling reason for females to retain the intron, it is puzzling why males, who lack SXL protein, splice this sequence from their *msl2* transcripts. Perhaps this sequence provides a subtle regulatory function not detected by our assays.

The finding that partial repression of *msl2* transcripts occurs if only a subset of poly(U) clusters is present lends support to our proposal that SXL mediates a second dosage compensation pathway in females^{4,20}. We proposed that SXL directly binds poly(U) clusters found in the 3' UTRs of a subset of X-linked transcripts such as *runt* to halve their translation. Partial repression by SXL is also consistent with the suggestion that MSL1 expression may be modulated through poly(U) clusters in the 3' UTR^{15,16}. Our results demonstrate that SXL has a function distinct from mRNA splicing regulation, and suggest that SXL can repress translation of transcripts to varying degrees, based on the number and location of binding sites in the target RNA.

Methods

MSL2 measurements in transgenic animals. Mutations were constructed by PCR with mutant primers²³, sequenced and substituted as *NotI*-*BglII* fragments into M2NG11. Transgenic animals were constructed²⁴ using the *w*⁺ marker in the pCasPeR vector²¹. Insertions were mapped to chromosomes, crossed into a *w*; *msl2*¹ *cn* background, assayed for rescue of mutant males, and examined by genomic Southern blots to determine which lines carried only a single copy of the transgene. Polytene chromosomes were prepared²⁵ from transgenic *msl2* larvae carrying a single copy of the transgene and stained with Hoechst 33258 and affinity-purified rabbit polyclonal antibodies to MSL2⁴. Four independent insertion sites were stained for each construct. Western blots were carried out as described¹⁶ using protein from one adult per lane. MSL2 was visualized with an affinity-purified polyclonal rabbit anti-MSL2 and an alkaline phosphatase-coupled secondary antibody.

RNA analysis. RT-PCR analysis was as described⁴ using Oligo number 75 (5'-AGTAGGGATCCACGAGC) and SuperScript II (Difco BRL) at 42°C to reverse-transcribe 3 µg DNase I-treated poly(A)⁺ mRNA. Oligo number 75

pairs perfectly with the wild-type *msl2* sequence, but has a 1-bp mismatch with the mutant *msl2*¹ sequence present in all transgenic lines, such that the endogenous PCR product is under-represented. PCR amplification was carried out using primers 1104 (5'-GGTCACACCTATGCCGACTGCAGCTAG) and 301 (5'-GCCGTAGCTCGCCGAGCCCGAGTTCAG) for 30 cycles annealing at 65°C. The samples were cut with *SmaI* and size-fractionated on 1.0% agarose. Nuclease protection assays were done with the HybSpeed RPA kit (Ambion) according to the supplier's instructions. The *msl2* antisense riboprobe was made from a template carrying the SXB1-2 sequence, extending from the *BamHI* to *SacI* sites, +147 to -188 relative to the start codon using T7 RNA polymerase.

SXL gel shift assay. 5' UTR RNA was transcribed by T7 RNA polymerase from an *SspI* to *SacI* fragment subcloned into pBluescript, linearized with *SacI* and made blunt-ended by T4 DNA polymerase. 3' UTR RNA was transcribed by T3 RNA polymerase from an *EcoRV* to 3' end subclone, linearized at the vector *BamHI* site. Plasmids containing wild-type or mutant 5' UTRs were linearized with *SacI*, blunt-ended with T4 DNA polymerase, and transcribed by T7 RNA polymerase. Production of GST-SXL and RNA gel shifts have been as described previously²², except that electrophoresis was at 200 volts for 4 h at room temperature.

Received 17 February; accepted 11 March 1997.

- Belote, J. M. & Lucchesi, J. C. Control of X chromosome transcription by the *maleless* gene in *Drosophila*. *Nature* **285**, 573–575 (1980).
- Baker, B. S., Gorman, M. & Marin, I. I. Dosage compensation in *Drosophila*. *Annu. Rev. Genet.* **28**, 491–521 (1994).
- Hilfiker, A., Hilfiker-Kleiner, D., Pannuti, A. & Lucchesi, J. C. *mof*, a putative acetyl transferase gene related to the Tip60 and MOZ human genes and to the SAS genes of yeast, is required for dosage compensation in *Drosophila*. *EMBO J.* **16**, 2054–2060 (1997).
- Kelley, R. L. *et al.* Expression of Msl-2 causes assembly of dosage compensation regulators on the X chromosomes and female lethality in *Drosophila*. *Cell* **81**, 867–877 (1995).
- Cline, T. W. Autoregulatory functioning of a *Drosophila* gene product that establishes and maintains the sexually determined state. *Genetics* **107**, 231–277 (1984).
- Lucchesi, J. C. & Skripsky, T. The link between dosage compensation and sex differentiation in *Drosophila melanogaster*. *Chromosoma* **82**, 217–227 (1981).
- Zhou, S. *et al.* Male-specific-lethal 2, a dosage compensation gene of *Drosophila* that undergoes sex-specific regulation and encodes a protein with a RING finger and a metallothionein-like cluster. *EMBO J.* **14**, 2884–2895 (1995).
- Bashaw, G. J. & Baker, B. S. The *msl2* dosage compensation gene of *Drosophila* encodes a putative DNA-binding protein whose expression is sex specifically regulated by *Sex-lethal*. *Development* **121**, 3245–3258 (1995).
- Bell, L. R., Horabin, J. I., Schedl, P. & Cline, T. W. Positive autoregulation of *Sex-lethal* by alternative splicing maintains the female determined state in *Drosophila*. *Cell* **66**, 229–239 (1991).
- Sosnowski, B. A., Belote, J. M. & McKeown, M. Sex-specific alternative splicing of RNA from the transformer gene results from sequence-dependent splice site blockage. *Cell* **58**, 449–459 (1989).
- Samuels, M. E. *et al.* RNA binding by Sxl proteins *in vitro* and *in vivo*. *Mol. Cell. Biol.* **14**, 4975–4990 (1994).
- Kozak, M. Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J. Biol. Chem.* **266**, 19867–19870 (1991).
- Geballe, A. P. & Morris, D. R. Initiation codons within 5'-leaders of mRNAs as regulators of translation. *Trends Biochem. Sci.* **19**, 159–164 (1994).
- Oh, S.-K., Scott, M. P. & Sarnow, P. Homeotic gene Antennapedia mRNA contains 5'-noncoding sequences that confer translational initiation by internal ribosome binding. *Genes Dev.* **6**, 1643–1653 (1992).
- Palmer, M. J. *et al.* The male specific lethal-one gene encodes a novel protein that associates with the male X chromosome in *Drosophila*. *Genetics* **134**, 545–557 (1993).
- Palmer, M. J., Richman, R., Richter, L. & Kuroda, M. I. Sex-specific regulation of the male-specific lethal-1 dosage compensation gene in *Drosophila*. *Genes Dev.* **8**, 698–706 (1994).
- Curtis, D., Lehmann, R. & Zamore, P. D. Translational regulation in development. *Cell* **81**, 171–178 (1995).

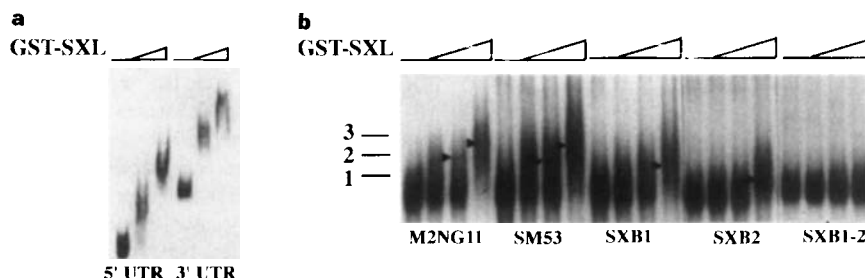


Figure 4 SXL binding to *msl2* RNA is reduced by mutations in the poly(U) stretches. **a**, Band-shift assays testing GST-SXL binding to a 246-nt region of the wild-type 5' UTR and a ~600-nt region of the wild-type 3' UTR. The 5' UTR RNA contains three potential binding sites, as the U16 stretch should comprise a double site to which two SXL molecules could bind cooperatively²². The 3' UTR RNA contains at least four separate binding sites. The first lane in each set contains RNA without SXL protein. **b**, Band-shift assays on the wild-type and

mutant RNA substrates shown in Fig. 1. For each 340-nt RNA substrate, the first lane lacks GST-SXL protein; in lanes 2 to 4, the protein concentration is increased in 1.5-fold increments. Numbers indicate three RNA-protein complexes of different sizes, which hypothetically contain 1, 2 or 3 molecules of SXL protein on a single RNA. M2NG11 and SM53 form complexes 2 and 3; SXB1, with an altered single site, forms complex 2; SXB2, with an altered double site, forms only complex 1; and SXB1-2, with no poly(U) stretches, forms no complex.

18. Caughman, S. W., Hentze, M. W., Rouault, T. A., Harford, J. B. & Klausner, R. D. The iron-responsive element is the single element responsible for iron-dependent translational regulation of ferritin biosynthesis. Evidence for function as the binding site for a translational repressor. *J. Biol. Chem.* **263**, 19048–19052 (1988).
19. Tarun, S. Z. & Sachs, A. B. Association of the yeast poly(A) tail binding protein with translation initiation factor eIF-4G. *EMBO J.* **15**, 7168–7177 (1996).
20. Kelley, R. L. & Kuroda, M. I. Equality for X chromosomes. *Science* **270**, 1607–1610 (1995).
21. Pirrotta, V. in *Vectors: A Survey of Molecular Cloning Vectors and Their Uses* (eds Rodriguez, R. L. & Denhardt, D. T.) 437–456 (Butterworths, Boston, 1988).
22. Wang, J. & Bell, L. R. The Sex-lethal amino terminus mediates cooperative interactions in RNA binding and is essential for splicing regulation. *Genes Dev.* **8**, 2072–2085 (1994).
23. Higuchi, R., Krummel, B. & Saiki, R. K. A general method of *in vitro* preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. *Nucleic Acids Res.* **16**, 7351–7367 (1988).
24. Rubin, G. M. & Spradling, A. C. Genetic transformation of *Drosophila* with transposable element vectors. *Science* **218**, 348–353 (1982).
25. Bone, J. R. *et al.* Acetylated histone H4 on the male X chromosome is associated with dosage compensation in *Drosophila*. *Genes Dev.* **8**, 96–104 (1994).

Acknowledgements. We thank M. J. Palmer, K. Chang and K. Copps for critical reading of the manuscript and K. Chang for the *msl1⁺* transgenic fly stock. This work was supported by grants from the NIH and NSF. M.I.K. is an Associate investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to M.I.K. (e-mail: mkuroda@bcm.tmc.edu).

Homotypic vacuolar fusion mediated by t- and v-SNAREs

Benjamin J. Nichols^{*}, Christian Ungermann[†], Hugh R. B. Pelham^{*}, William T. Wickner[‡] & Albert Haas^{††}

^{*} Medical Research Council, Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

[†] Department of Biochemistry, Dartmouth Medical School, Hanover, New Hampshire 03755-3844, USA

[‡] Present address: Lehrstuhl für Mikrobiologie, Theodor-Boveri-Institut für Biowissenschaften, Universität Würzburg, Am Hubland, 97074 Würzburg, Germany

Membrane fusion is necessary both in the eukaryotic secretory pathway and for the inheritance of organelles during the cell cycle. In the secretory pathway, heterotypic fusion takes place between small transport vesicles and organelles. It requires N-ethylmaleimide-sensitive fusion protein (NSF/Sec18p), soluble NSF attachment proteins (SNAPs/Sec17p) and SNAP receptors (SNAREs). SNAREs are integral membrane proteins (v-SNAREs on vesicles, t-SNAREs on the target organelles) and are thought to provide specificity to the fusion process^{1–5}. It has been suggested that Sec17p and Sec18p bind to v-SNARE/t-SNARE complexes and mediate the membrane fusion event^{1–3}. Homotypic fusion of yeast

vacuoles also requires Sec17p and Sec18p (ref. 6), but *in vitro* they are needed only to 'prime' the vacuoles, not for subsequent docking or fusion^{7,8}. It has been unclear whether these reactions involve SNAREs that are similar to those previously identified in heterotypic fusion systems and, hence, whether the actions of Sec18p/NSF and Sec17p/αSNAP in these systems can be compared. Here we identify typical v- and t-SNAREs on the yeast vacuolar membrane. Although both are normally present, vacuoles containing only the v-SNARE can fuse with those containing only the t-SNARE. Vacuoles containing neither SNARE cannot fuse with those containing both, demonstrating that docking is mediated by cognate SNAREs on the two organelle membranes. Even when t- and v-SNAREs are on separate membranes, Sec17p and Sec18p act at the priming stage. Their action is not required at the point of assembly of the SNARE complex, nor for the fusion event itself.

We surveyed the yeast genome database for proteins related to the syntaxin family of t-SNAREs and the VAMP/synaptobrevin family of v-SNAREs^{9,10}, and the candidates included the open reading frames (ORFs) YOR106w (*VAM3*) and YLR093c (*NYV1*). Both have typical SNARE features, including a carboxy-terminal membrane anchor preceded by a region that has the potential to form an α-helical coiled-coil. The coiled-coil region of YOR106w shows between 40 and 18% identity to the yeast t-SNAREs Pep12p (ref. 11), Sed5p (ref. 4), Sso1p (ref. 9) and Ufe1p (ref. 12), whereas the corresponding region of YLR093c is 41% identical to Snc1p, a v-SNARE involved in transport to the plasma membrane¹⁰ (Fig. 1a). Immunofluorescence microscopy of the functional, epitope-tagged proteins expressed in yeast showed that each colocalizes with a subunit of the vacuolar ATPase and hence is located on the vacuolar membranes (Fig. 1b). Furthermore, co-immunoprecipitation of the two proteins could be observed, which suggests that they are capable of interacting with each other (data not shown).

Deletion of YOR106w was not lethal, but resulted in abnormal vacuoles. In the mutant, endocytosis of the membrane dye FM4-64 (ref. 13) led to labelling of numerous small organelles (Fig. 2a), as did staining with antibodies specific for the vacuolar ATPase (not shown). The mutant cells did not secrete significant amounts of the vacuolar protein carboxypeptidase Y (CPY)¹⁴ but they contained unusually high levels of CPY precursors (Fig. 2b). This phenotype is very similar to that of *ypt7* (ref. 15), a Rab-like GTPase required for vacuolar fusion¹⁶, and of the *vam* mutants¹⁷. While this work was in progress, a GenBank entry identified YOR106w as *VAM3* (Y. Wada, Y. Ohsumi and A. Hirata, accession number U57827). Purification of the organelles¹⁸ from the *vam3* deletion strain demonstrated that they contain roughly normal amounts of mature CPY and the other

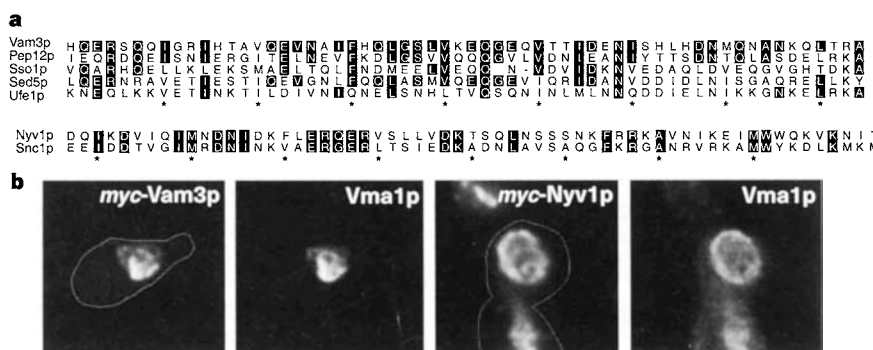


Figure 1 *VAM3* and *NYV1* encode SNARE-like proteins located in the vacuole. **a**, Sequences of potential coiled-coil domains from Vam3p and Nyv1p. Residues 193–253 of Vam3p are aligned with other t-SNAREs: residues 198–258 of Pep12p, 252–312 of Sed5p, 192–252 of Sso1p and 258–318 of Ufe1p. Residues 171–231 of Nyv1p are aligned with residues 34–94 of the v-SNARE Snc1p. Heptad repeats of hydrophobic amino acids are indicated (*). Residues showing identity with Vam3p

or Nyv1p are highlighted. **b**, Left- and right-hand images correspond to double labelling of cells containing Myc-tagged Vam3p or Nyv1p, with antibodies against the Myc-epitope and against the vacuolar membrane protein Vma1p, visualized by confocal microscopy. Only the vacuoles are visible at this exposure, but the outlines of the cells have been traced in the left-hand panels.

vacuolar proteins alkaline phosphatase (Pho8p), vacuolar ATPase (the Vma1p subunit) and proteinase A (Pep4p) (Fig. 2c). Because they lack a presumptive vacuolar t-SNARE, these organelles may form by an abnormal process, such as endosome maturation. Nevertheless, their content indicates that they are functionally vacuoles.

Deletion of YLR093c, which we have named *NYV1* (new yeast v-SNARE) had not effect on viability. $\Delta nyv1$ cells tend to have a single large vacuole (Fig. 2a). Processing of CPY appeared normal in $\Delta nyv1$ cells (Fig. 2b), and vacuoles purified from these cells contained normal levels of vacuolar proteins (Fig. 2c).

The presence of typical t- and v-SNAREs on vacuolar membranes, together with previous biochemical data, suggests the working model for vacuolar fusion depicted in Fig. 3. In this model, the individual SNAREs are activated in an ATP-dependent manner by Sec17p and Sec18p. Sec17p is released and the cytosolic factor LMA1 preserves the vacuoles in their activated state^{8,19}. Docking requires Ypt7p (ref. 10) and is mediated by the formation of t-SNARE/v-SNARE complexes that link the vacuoles, allowing subsequent fusion.

This model predicts that both v-SNARE (Nyv1p) and t-SNARE (Vam3p) will be required for efficient fusion, but that it might suffice for each vacuole to contain only one of them. To test this, we deleted the *VAM3* and *NYV1* genes individually in two strains, one lacking the vacuolar alkaline phosphatase Pho8p, and the other lacking Pep4p, the protease that processes and thereby activates proPho8p. Fusion of purified vacuoles from these strains would lead to mixing of vacuole contents and hence production of phosphatase activity, which can be assayed spectrophotometrically^{18,21} (Fig. 3).

The results shown in Fig. 4a provide strong support for this model. Removal of the v-SNARE Nyv1p from one of the fusion partners had relatively little effect on the fusion reaction, but when it was absent from both the efficiency dropped dramatically. Similarly, complete removal of the t-SNARE Vam3p virtually abolished fusion, whereas when it was present on only one partner, fusion could still proceed, albeit at reduced efficiency compared to the normal reaction. The low signal in this case may mean that optimal fusion requires the t-SNARE to be present on both fusion partners, but it may also be a consequence of the smaller size of the Vam3p-deficient vacuoles, which leads to smaller amounts of Pep4p and Pho8p being mixed per fusion event than is the case with normal vacuoles. Nevertheless, fusion clearly occurred when Vam3p was present only on one partner and Nyv1p only on the other, indicating that a heterotypic v-SNARE/t-SNARE interaction is sufficient to drive fusion.

We next asked whether fusion could occur between vacuoles containing both the v- and t-SNAREs and vacuoles containing neither. We prepared double-deletion strains lacking Nyv1p and Vam3p and found that they contained vacuoles similar in morphology to those in the *vam3*-deleted strains, with similar levels of CPY, proteinase A and pro-alkaline phosphatase (data not shown). However, they were completely unable to fuse with normal vacuoles (Fig. 4b). This result strongly suggests that fusion requires the formation of a SNARE complex that bridges the two fusing membranes; the presence of v- and t-SNAREs on one membrane is not sufficient. It also shows that the fusion signal measured in our assays is entirely dependent on the SNAREs we have identified—no activity ascribable to other SNAREs or to alternative fusion mechanisms could be detected.

Because the Vam3p-deficient strain has small vacuoles, we used antibody inhibition of fusion to provide independent confirmation of the role of this protein. Fab fragments prepared from anti-Vam3p IgG totally inhibited a fusion reaction containing parental strain vacuoles, whereas Fab fragments prepared from preimmune serum IgG did not (Fig. 4c, d). The antibodies provide a tool to identify the stage in the overall reaction at which Vam3p is required. By varying the time of incubation before the addition of antibodies, we found

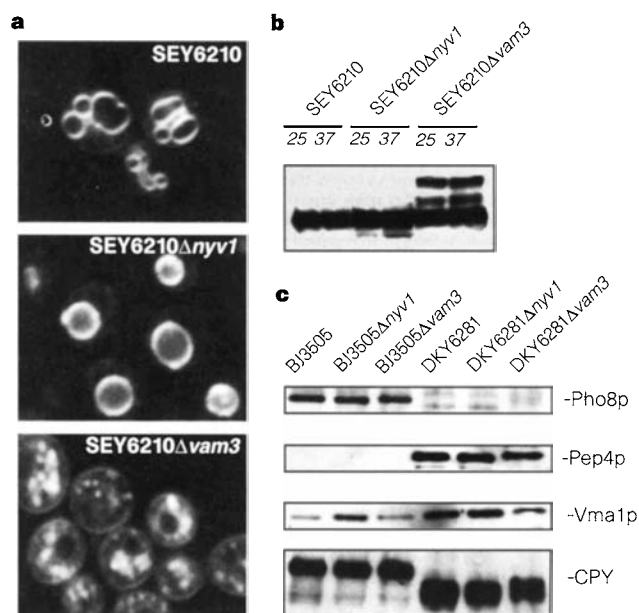


Figure 2 Phenotypes of *VAM3* and *NYV1* deletion strains. **a**, Vacuolar morphology. Vacuoles from the parental strain SEY6210 and from the two deletion strains were labelled using the fluorescent dye FM4-64 (ref. 13), and visualized by confocal microscopy. Cells were collected in log-phase growth. **b**, CPY processing. Cells were grown at 25°C, or shifted to 37°C for 2 h. CPY was detected by immunoblotting of whole-cell extracts. The mature, vacuolar form of CPY is labelled mCPY. The upper band of the CPY precursors accumulating in the *vam3* deletion strain is the size of P2 CPY¹⁴. The lower band is smaller than P1 CPY, and probably corresponds to the abnormal intermediate seen in a *ypt7* mutant¹⁵. **c**, The effects of disrupting *VAM3* and *NYV1* on the levels of vacuolar marker proteins. Equivalent amounts of vacuolar protein, purified as described in Methods, were loaded in each lane, and the proteins indicated detected with appropriate antisera. Parental strains are BJ3505 and DKY6281. The vacuolar fusion assay depends on the absence of Pho8p in DKY6281 and of Pep4p in BJ3505. The increase in mass of CPY in BJ3505 is due to absence of proteolysis by Pep4p¹⁴.

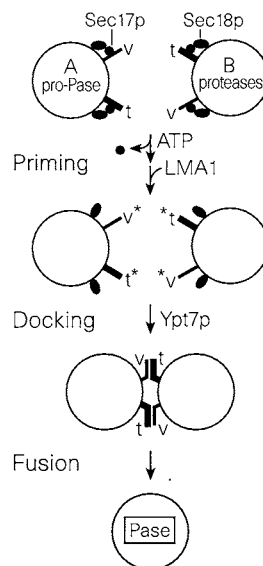


Figure 3 Model for vacuolar fusion (see text for details). The *in vitro* assay measures fusion between vacuoles containing the Pho8p pro-alkaline phosphatase (Pase) but lacking the Pep4p protease (from strain A, BJ3505) and those containing the protease but not the phosphatase (from strain B, DKY6281). The v- and t-SNAREs (v and t) correspond to Nyv1p and Vam3p, respectively. Asterisks indicate 'primed' SNAREs.

that the fusion reaction became resistant to anti-Vam3p Fabs with identical kinetics to the acquisition of resistance to Ypt7p antibodies (Fig. 4d). Ypt7p acts at the point of vacuole docking, as identified by determining when fusion becomes resistant to dilution²⁰. This occurs after the requirement for Sec17p and well before the actual fusion event itself, which confers resistance to chilling (Fig. 4d). These data implicate Vam3p in the docking step of the overall fusion process, as predicted by our model.

To investigate further the point in the fusion reaction at which Vam3p is required, we preincubated the parental-strain vacuoles in separate tubes under reaction conditions, then mixed them and measured their fusion. Fusion progressed with high efficiency when the vacuoles were preincubated briefly, but its efficiency decreased with increasing time of separate incubation (Fig. 4e), as previously described^{8,19,20}. Addition of antibodies at the time of mixing showed that Sec17p was no longer required once the vacuoles had been preincubated separately for 5–10 min⁸. However, anti-Vam3p, like anti-Ypt7p, totally blocked fusion if added at the time of mixing, regardless of the preincubation time (Fig. 4e). This clearly indicates that docking is required for fusion to become resistant to Vam3p antibodies, consistent with a direct role for Vam3p in docking.

We next addressed the role of Sec17p and Sec18p. We have previously shown that these act before docking of wild-type vacuoles⁸. One possibility would be that, during homotypic fusion, they serve to form or disrupt v-SNARE/t-SNARE complexes

on each vacuole before the assembly of the docking complex. If so, then the role of Sec17p and Sec18p might be different when the vacuoles each possess only a single SNARE. However, fusion of Nyv1p-containing vacuoles with Vam3p-containing ones was still sensitive to anti-Sec17p and anti-Sec18p antibodies, being inhibited by up to 92%. Furthermore, addition of antibodies to the fusion reaction at various times shows that, as with wild-type vacuoles, the requirement for these proteins could be fulfilled before the docking step (measured in this experiment by sensitivity to GDI, which prevents Ypt7p from mediating docking¹⁶); compare Fig 4d with f. We conclude that Sec17p and Sec18p can act on single SNAREs, and that although they are required to activate the SNAREs for docking, they are not themselves needed for docking or fusion.

Our results provide strong support for the model depicted in Fig. 3 and for the SNARE hypothesis in general. However, they show that Sec18p and Sec17p prime the SNAREs for docking, rather than acting at the point of fusion. One additional caveat concerns the role of the v-SNARE Nyv1p. Removal of this SNARE does not greatly affect vacuole formation *in vivo*, and does not completely abolish fusion *in vitro*. The residual fusion is completely dependent on Vam3p, and we suggest that it is mediated by interactions between Vam3p molecules on different membranes. Other t-SNAREs such as Sed5p (ref. 22) and syntaxin²³ are capable of self-association, and Vam3p may do the same. Interestingly, removal of Snc1p and Snc2p, the only v-SNAREs so far implicated in transport to the yeast plasma

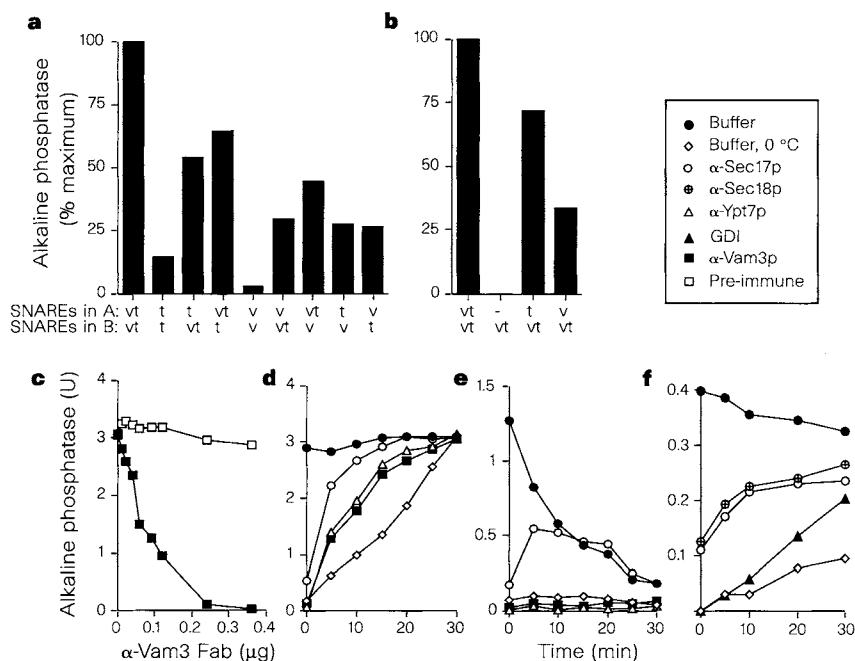


Figure 4 Vacuolar fusion assay. **a, b**, Fusion of a Δ vam3, Δ nyv1 and parental strain vacuoles. Alkaline phosphatase activity is expressed as a percentage of that obtained by fusion of vacuoles from the parental strains¹⁸, which was 1.4 units in experiment A and 0.65 units in experiment B. The t- and v-SNAREs present in the Δ pep4 strain (strain A, BJ3505) and the Δ pho8 strain (strain B, DKY6281) are indicated (t, Vam3p; v, Nyv1p). Fusion-independent phosphatase activity measured in parallel reactions incubated without ATP, or on ice (0.17–0.29 units) has been subtracted from the data shown. **c**, Vacuolar fusion in the presence of anti-Vam3p Fab fragments. Reactions were performed in the presence of the indicated amount of Fab fragments prepared from preimmune or anti-Vam3p IgGs. Fusion-independent phosphatase activity of 0.20 units was subtracted. **d**, Kinetics of acquisition of resistance to anti-Sec17p, anti-Ypt7p and anti-Vam3p antibodies. A fusion reaction containing each of the two types of parental strain vacuoles was incubated at 25 °C. At various times, 30-μl aliquots were mixed with appropriate antibodies, or a buffer-only control, as indicated. Additional samples were transferred to ice. Alkaline phosphatase activity was measured in all

samples at the end of the experiment. Fusion-independent phosphatase activity of 0.19 units was subtracted. **e**, Requirement for Sec17p, but not Vam3p or Ypt7p, can be satisfied while vacuoles are in separate tubes. This experiment was carried out in parallel with **d**, and identical amounts of antibodies were used. Symbols used are also the same. Aliquots (15 μl) from two ongoing separate reactions, containing only DKY6281 or BJ3505 vacuoles, were mixed at the time-points indicated into tubes that contained 3 μl of either the antibodies or reaction buffer. Each mix was left at the reaction temperature for 70 min, then alkaline phosphatase activity was determined. Fusion-independent phosphatase activity of 0.19 units was subtracted. **f**, Sec17p and Sec18p act before docking when t- and v-SNAREs are on separate membranes. This experiment was similar to that in **d**, except that the vacuoles from strain A contained only Vam3p and those from strain B only Nyv1p. GDI¹⁶ was added at 66 μg ml⁻¹. Note that fusion took longer to reach completion than with parental strain vacuoles (as shown in **d**), possibly because of the small size of the Vam3p-deficient vacuoles.

membrane, does not completely abolish secretion¹⁰ although removal of the corresponding t-SNAREs (Sso1p and Sso2p) is lethal⁹.

The homotypic vacuolar fusion system, which possesses many of the features of heterotypic systems, should prove a useful model for further biochemical analysis. Moreover there is increasing evidence that both t-SNAREs and v-SNAREs are present in synaptic²⁴, endocytic²⁵, and endoplasmic reticulum-derived vesicles^{22,26}, and it may be that fusion of such vesicles to each other or to their targets can occur by a mechanism very similar to that of vacuolar fusion. □

Methods

Yeast strains. The *Saccharomyces cerevisiae* strain used for immunofluorescence and characterization of deletion phenotypes was SEY6210 (*MAT α leu2-3, -112 ura3-52, his3- Δ 200, trp1- Δ 901, lys2-801, suc2- Δ 9). Vacuolar fusion assay used BJ3505 (*MAT α pep4::HIS3 prb1- Δ 1.6R his3 lys2-208 trp1- Δ 101 ura3-52 gal2 can*) and DKY6281 (*MAT α leu2-3, -112 ura3-52 his3- Δ 200 trp1- Δ 901 lys2-801 suc2- Δ 9 pho8::TRP1*). *VAM3* and *NYV1* ORFs were completely replaced by homologous recombination with *HIS5* from *Schizosaccharomyces pombe*²⁷ or with *TRP1* or *URA3*, using PCR-generated flanking homologous regions²⁷.*

Immunolocalization. PCR-amplified *NYV1* and *VAM3* were expressed from the multicopy, TPI promoter plasmid JS209 (ref. 28), which introduces an amino-terminal Myc epitope, and detected by immunofluorescence²⁹.

Antibodies. Fluorophore-conjugated antibodies were purchased from Santa Cruz Biotech and Sigma. Pep4p, Pho8p, Sec17p and Ypt7p antibodies were produced as described^{6,16}. An antiserum was raised in rabbits against a His₆-tagged fusion protein (produced using pTrcHis-A, Invitrogen) containing amino acids 1–253 of Vam3p. For the experiments shown in Fig. 4, antibodies were used as follows: anti-Vam3p Fabs (from total immune IgG) at 10.5 μ g ml⁻¹ per reaction, anti-Ypt7p (from total immune IgG) at 215 μ g ml⁻¹, anti-Sec17p (affinity-purified antibodies) at 60 μ g ml⁻¹.

Vacuolar purification and fusion assay. Methods used to produce highly purified vacuoles and to assay fusion have been described¹⁸, and the reaction buffer was as described⁸. Cytosol for addition to fusion reactions was prepared from a strain deficient in alkaline phosphatase activity (K91-1A) using French-press lysis of yeast spheroplasts⁶.

Received 14 February; accepted 12 March 1997.

- Rothman, J. E. Mechanisms of intracellular protein transport. *Nature* **372**, 55–63 (1994).
- Söllner, T., Bennett, M. K., Whiteheart, S., Scheller, R. H. & Rothman, J. E. A protein assembly-disassembly pathway *in vitro* that may correspond to sequential steps of synaptic vesicle docking, activation, and fusion. *Cell* **75**, 409–418 (1993).
- Söllner, T. et al. SNAP receptors implicated in vesicle targeting and fusion. *Nature* **362**, 318–324 (1993).
- Hardwick, K. & Pelham, H. R. B. SED5 encodes a 39-kD integral membrane protein required for vesicular transport between the ER and the Golgi complex. *J. Cell Biol.* **119**, 513–521 (1992).
- Sogaard, M. et al. A rab protein is required for the assembly of SNARE complexes in the docking of transport vesicles. *Cell* **78**, 937–948 (1994).
- Haas, A. & Wickner, W. Homotypic vacuole fusion requires Sec17p (Yeast α -SNAP) and Sec18p (yeast NSF). *EMBO J.* **15**, 3296–3305 (1996).
- Conradt, B., Haas, A. & Wickner, W. Determination of four biochemically distinct, sequential stages during vacuole inheritance *in vitro*. *J. Cell Biol.* **126**, 99–110 (1994).
- Mayer, A., Wickner, W. & Haas, A. Sec18p (NSF)-driven release of Sec17p (α -SNAP) can precede docking and fusion of yeast vacuoles. *Cell* **85**, 83–94 (1996).
- Aalto, M. K., Ronne, H. & Keranen, S. Yeast syntaxins Sso1p and Sso2p belong to a family of related membrane proteins that function in vesicular transport. *EMBO J.* **12**, 4095–4104 (1993).
- Protopopov, V., Govindan, B., Novick, P. & Gerst, J. E. Homologs of the synaptobrevin/VAMP family of synaptic vesicle proteins function on the late secretory pathway in *S. cerevisiae*. *Cell* **74**, 855–861 (1993).
- Becherer, K. A., Rieder, S. E., Emr, S. D. & Jones, E. W. Novel syntaxin homologue, Pep12p, required for the sorting of luminal hydrolases to the lysosome-like vacuole in yeast. *Mol. Cell Biol.* **7**, 579–594 (1996).
- Lewis, M. J. & Pelham, H. R. B. SNARE-mediated retrograde traffic from the Golgi complex to the endoplasmic reticulum. *Cell* **85**, 205–215 (1996).
- Vida, T. A. & Emr, S. D. A new vital stain for visualizing vacuolar membrane dynamics and endocytosis in yeast. *J. Cell Biol.* **128**, 779–792 (1995).
- Stevens, T. H., Esmon, B. & Schekman, R. Early stages in the yeast secretory pathway are required for transport of carboxypeptidase Y to the vacuole. *Cell* **30**, 439–448 (1982).
- Wichmann, H., Hegst, L. & Gallwitz, D. Endocytosis in yeast: evidence for the involvement of a small GTP-binding protein (Ypt7p). *Cell* **71**, 1131–1142 (1992).
- Haas, A., Schlegelmann, D., Lazar, T., Gallwitz, D. & Wickner, W. The GTPase Ypt7p of *Saccharomyces cerevisiae* is required on both partner vacuoles for the homotypic fusion step of vacuole inheritance. *EMBO J.* **14**, 5258–5270 (1995).
- Wada, Y. & Anraku, Y. Genes for directing vacuolar morphogenesis in *Saccharomyces cerevisiae*. I. Isolation and characterization of two classes of vam mutants. *J. Biol. Chem.* **267**, 18671–18675 (1992).
- Haas, A. A quantitative assay to measure homotypic vacuole fusion *in vitro*. *Meth. Cell Sci.* **17**, 283–294 (1995).

- Xu, Z., Mayer, A., Muller, E. & Wickner, W. A heterodimer of thioredoxin and I82 cooperates with Sec18p (NSF) to promote yeast vacuole inheritance. *J. Cell Biol.* **136**, 299–306 (1997).
- Mayer, A. & Wickner, W. Docking of yeast vacuoles is catalysed by the Ras-like GTPase Ypt7p after symmetric priming by Sec18p (NSF). *J. Cell Biol.* **136**, 307–317 (1997).
- Haas, A., Conradt, B. & Wickner, W. I. G-protein ligands inhibit *in vitro* reactions of vacuole inheritance. *J. Cell Biol.* **126**, 87–97 (1994).
- Banfield, D. K. et al. Localization of Sed5, a putative vesicle targeting molecule, to the *cis*-Golgi network involves both its transmembrane and cytoplasmic domains. *J. Cell Biol.* **127**, 357–371 (1994).
- Yoshida, A. et al. HPC-1 is associated with synaptotagmin and ω -conotoxin receptor. *J. Biol. Chem.* **267**, 24925–24928 (1992).
- Walch-Solimena, C. et al. The t-SNAREs syntaxin 1 and SNAP25 are present on organelles that participate in synaptic vesicle recycling. *J. Cell Biol.* **128**, 637–645 (1995).
- Steel, G. J., Tagaya, M. & Woodman, P. G. Association of the fusion protein NSF with clathrin-coated vesicle membranes. *EMBO J.* **15**, 745–752 (1996).
- Dascher, C., Matteson, J. & Balch, W. E. Syntaxin-5 regulates endoplasmic-reticulum to Golgi transport. *J. Biol. Chem.* **269**, 29363–29366 (1994).
- Wach, A., Brachat, A., Poehlmann, R. & Phillipsen, P. New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. *Yeast* **10**, 1793–1808 (1994).
- Semenza, J. C., Hardwick, K. G., Dean, N. & Pelham, H. R. B. ERD2, a yeast gene required for the receptor-mediated retrieval of luminal ER proteins from the secretory pathway. *Cell* **61**, 1349–1357 (1990).
- Kilmartin, J. V. & Adams, A. E. M. Structural rearrangements of tubulin and actin during the cell cycle of the yeast *Saccharomyces*. *J. Cell Biol.* **98**, 922–933 (1984).

Acknowledgements. We thank our colleagues M. Lewis and J. Rayner for advice and reagents, S. Emr and T. Stevens for helpful discussions, and S. Dhruvakumar for assistance with immunolocalization studies on Vam3p. This work was supported in part by a grant to W.T.W. from the NIH and by the Deutsche Forschungsgemeinschaft (C.U.).

Correspondence and requests for materials should be addressed to H.R.B.P. (e-mail: hp@mrc-lmb.cam.ac.uk).

Regulation of gene expression by small molecules

Joel M. Gottesfeld*, Laura Neely*, John W. Trauger†, Eldon E. Baird† & Peter B. Dervan†

* Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037, USA

† Division of Chemistry and Chemical Engineering and Beckman Institute, California Institute of Technology, Pasadena, California 91125, USA

Small molecules that target specific DNA sequences have the potential to control gene expression. Ligands designed for therapeutic application must bind any predetermined DNA sequence with high affinity and permeate living cells. Synthetic polyamides containing *N*-methylimidazole and *N*-methylpyrrole amino acids have an affinity and specificity for DNA comparable to naturally occurring DNA-binding proteins¹. We report here that an eight-ring polyamide targeted to a specific region of the transcription factor TFIID binding site interferes with 5S RNA gene expression in *Xenopus* kidney cells. Our results indicate that pyrrole-imidazole polyamides are cell-permeable and can inhibit the transcription of specific genes.

Two approaches for the development of synthetic transcriptional antagonists have been reported^{2–7}. Oligodeoxyribonucleotides that recognize the major groove of double-helical DNA through triple-helix formation bind to a broad range of sequences with high affinity and specificity^{2,3}. Although oligonucleotides and their analogues interfere with gene expression^{4,5}, the triple-helix approach is limited to purine tracks and suffers from poor cellular uptake. There are a few examples of cell-permeable carbohydrate-based ligands that interfere with transcription factor function^{6,7}, but oligosaccharides cannot currently recognize a broad range of DNA sequences.

Pyrrole-imidazole polyamides represent the only class of small molecules so far that can bind any predetermined DNA sequence. DNA recognition depends on side-by-side amino-acid pairings in the minor groove^{8,9}. A pairing of imidazole (Im) opposite pyrrole (Py) targets a G-C base pair, and Py/Im targets a C-G base pair^{8,9}. A Py/Py combination is degenerate and targets both T-A and A-T base pairs^{8–11}. The generality of these pairing rules has been demon-

strated by targeting a variety of sequences of 5–13 base pairs (bp)^{12–14} and is supported by NMR evidence¹⁵.

The relatively small number of genes transcribed by RNA polymerase III provides a simple system in which to explore the selectivity and efficiency of polyamides as regulators of gene expression. Well established methods exist for assessing in living cells the status of RNA polymerase III transcription complexes on the genes encoding the small 5S ribosomal RNA¹⁶. The role of the nine-zinc-finger protein TFIIIA in transcriptional regulation of the 5S RNA gene by RNA polymerase III has been characterized^{17,18}. Zinc-fingers 1–3, 5 and 7–9 bind the internal control region (ICR) of the gene through base-specific interactions in the major groove^{19–22}; fingers 4 and 6 are essential for high-affinity DNA binding²³ and bind in or across the minor groove^{21,22} (Fig. 1). The 6-base-pair (bp) DNA sequence 5'-AGTACT-3' is in the binding site for finger 4. According to the pairing rules, this sequence would be bound in a hairpin motif by the eight-ring polyamide **1** of sequence ImPyPyPy- γ -ImPyPyPy- β . Quantitative footprint titration experiments reveal that polyamide **1** selectively binds the 6-bp target sequence with a dissociation constant K_d of 0.03 nM¹, a higher affinity than that of TFIIIA for its 50-bp site ($K_d \approx 1$ nM)²³.

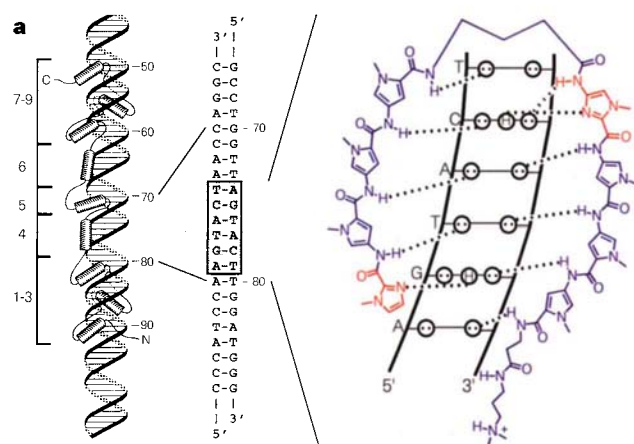
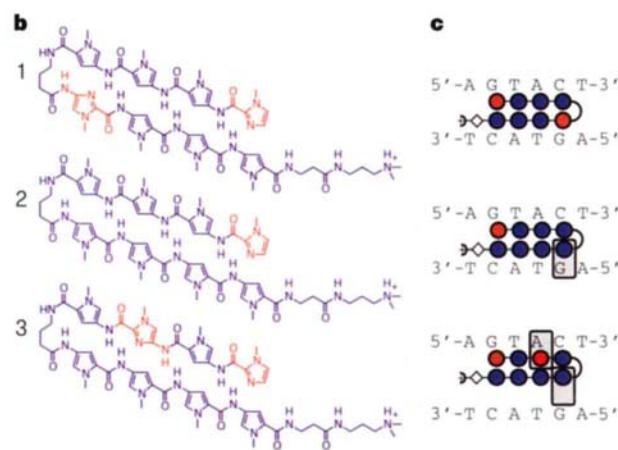


Figure 1 **a** Left, model of the nine-zinc-finger protein TFIIIA with the 5S RNA gene internal control region (ICR)^{21,22}; middle, sequence of the ICR recognized by zinc-finger 4 in the minor groove; right, complex of hairpin polyamide **1** with its target site, 5'-AGTACT-3'. Circles with dots represent lone pairs on N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of

Mismatch eight-ring polyamides **2** and **3** have 100-fold ($K_d = 2.0$ nM) and 1,000-fold ($K_d = 33$ nM) lower affinity, respectively, for the 5'-AGTACT-3' site (Fig. 1).

We examined the effect of polyamide **1** on TFIIIA binding to a restriction fragment isolated from a plasmid containing the 5S RNA gene. Zf1-3, a recombinant TFIIIA analogue lacking fingers 4–9, binds in the major groove of the C-block promoter element^{21,22} (Fig. 1). DNase I footprinting demonstrates that Zf1-3 and polyamide **1** can co-occupy the same DNA molecule. When 5 nM polyamide **1** was preincubated with the same DNA target, the binding of 9-finger TFIIIA was inhibited by >90% (Fig. 2). The differential inhibition of Zf1-3 and full-length TFIIIA is evidence that finger 4 interacts with or is placed in the minor groove. Polyamide **1** does not inhibit TFIIIA binding to 5S RNA.

Transcription of the 5S RNA gene *in vitro* was monitored in the presence of increasing concentrations (10–60 nM) of polyamide **1** (Fig. 3). In these experiments, polyamide **1** was added to a plasmid containing the 5S RNA gene before addition of exogenous TFIIIA (12 nM) and a crude extract derived from unfertilized *Xenopus* eggs. As a control, a tyrosine-transfer-RNA gene was included on a separate plasmid in these reactions. The tRNA gene has an upstream



guanine. **b**, Structures of polyamides ImPyPyPy- γ -ImPyPyPy- β -Dp (**1**), ImPyPyPy- γ -PyPyPyPy- β -Dp (**2**), and ImPyImPy- γ -PyPyPyPy- β -Dp (**3**). Dp, dimethylamino-propylamide. **c**, Binding models: red and blue circles represent imidazole and pyrrole rings, respectively, half-circles represent γ -aminobutyric acid, and diamonds represent β -alanine. Hydrogen-bond mismatches are highlighted.

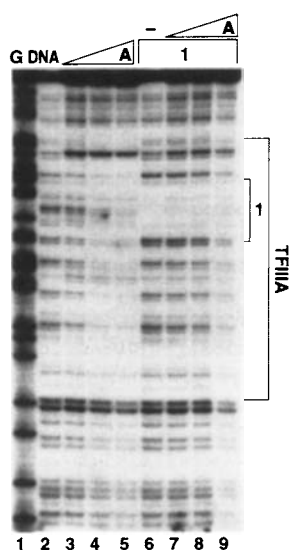


Figure 2 DNase I footprinting analysis of the binding of polyamide **1** and TFIIIA to the 5S RNA gene ICR. 5'-end-labelled restriction fragments were derived from the 5S RNA gene by standard methods and footprinting reactions were done as described²³. Lanes: 1, DMS G-specific sequencing ladder; 2, protein-free DNA; 3–5, digestion products obtained in the presence of 0.5 nM, 1 nM and 2 nM TFIIIA, respectively; 6, digestion products obtained in the presence of 5 nM of **1**; 7–9, digestion products of reactions in which the DNA was incubated with **1** for 30 min before addition of TFIIIA (at the same concentrations as in lanes 3–5, respectively) and incubated for an additional 30 min before DNase treatment.

binding site for **1**, but does not participate in a predicted protein–polyamide interaction²⁴. Both genes are actively transcribed in this system, either individually (lanes 1 and 15) or in mixed-template reactions (lane 2). Addition of 60 nM polyamide **1** inhibits 5S gene transcription by >80% (lane 6). Only a small degree of nonspecific inhibition of tRNA transcription is observed at the concentrations of polyamide **1** required for efficient 5S RNA inhibition. The targeted 5S RNA gene is inhibited ~10-fold more effectively than the control tRNA gene (Fig. 3). Mismatch polyamides **2** and **3** do not inhibit 5S RNA transcription at concentrations up to 60 nM (Fig. 3). If the TFIIIA–DNA complex is first allowed to form, 30 nM polyamide **1** added, and the mixture incubated for 90 min before adding egg extract, 5S RNA transcription is efficiently inhibited (80%). Shorter incubation times result in less inhibition. The necessary incubation time of 90 min is similar to the measured half-life of the TFIIIA–DNA complex²³ and supports

the idea that **1** forms a more stable complex with DNA than does TFIIIA.

The effect of the polyamides on transcription of the 5S RNA gene *in vivo* was tested. *Xenopus* kidney-derived fibroblasts were grown in the presence of increasing concentrations of polyamide **1** in the culture medium for various times. We found that concentrations of polyamide up to 1 μ M were not toxic, as measured by cell density, if growth was for less than 72 hours. Nuclei were prepared from cells by hypotonic lysis and equivalent amounts of the isolated nuclei from control and treated cells were used as templates for transcription with exogenous RNA polymerase III and labelled and unlabelled nucleoside triphosphates. This experiment monitors the occupancy of class III genes with active transcription complexes¹⁶. 5S RNA transcription can easily be assessed because the repetitive 5S genes give rise to a prominent band on a denaturing polyacrylamide gel. Figure 4a shows the autoradiogram of this experiment and Fig. 4b

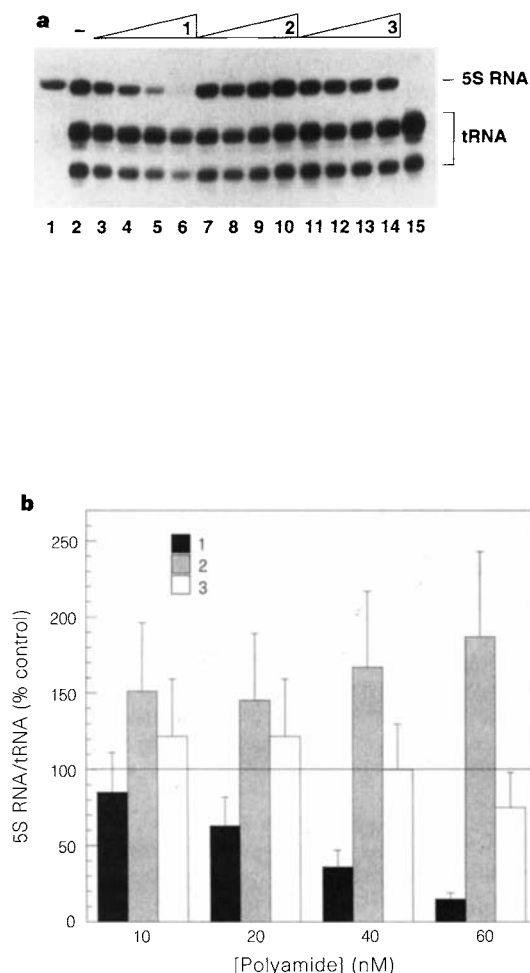


Figure 3 Inhibition of 5S RNA gene transcription *in vitro*. **a**, DNA templates were incubated with polyamide for 30 min before addition of TFIIIA, a cytoplasmic extract prepared from unfertilized *Xenopus* eggs, and labelled and unlabelled nucleoside triphosphates²⁶. Reactions contained either the somatic-type 5S RNA gene²⁷ (lane 1), a tRNA^{tyr} gene²⁸ (lane 15), or a mixture of both genes (lanes 2–14), and the following final concentrations of the indicated polyamide: 0 nM, lanes 1, 2, 15; 10 nM, lanes 3, 7, 11; 20 nM, lanes 4, 8, 12; 40 nM, lanes 5, 9, 13; 60 nM, lanes 6, 10, 14. Transcription reactions were for 1 h at ambient temperature and RNA products were analysed on a denaturing polyacrylamide gel. The positions of 5S RNA and tRNA primary and processed transcripts are indicated at the right of the autoradiogram. **b**, Graphic representation of inhibition results expressed as the ratio of 5S RNA to tRNA transcription relative to the ratio obtained in the absence of polyamide. Error bars represent the estimated standard deviation.

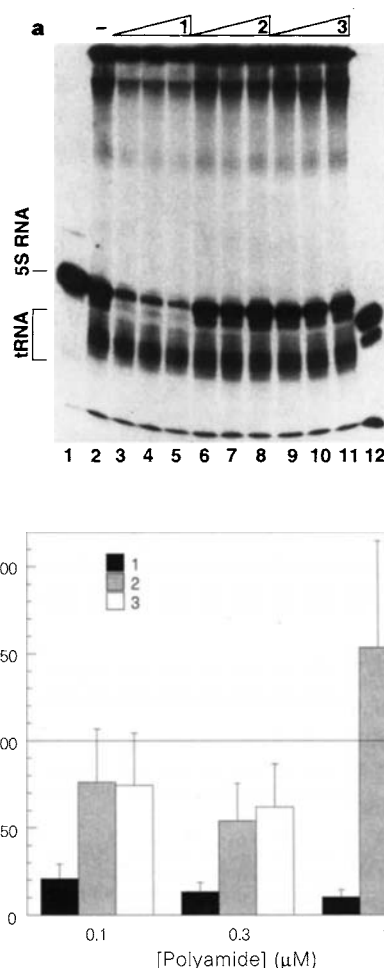


Figure 4 Inhibition of 5S RNA gene transcription complex formation *in vivo*. **a**, Nuclei were prepared from *Xenopus* kidney-derived fibroblasts grown in culture in the presence or absence of polyamides^{16,29}. Polyamides were included in culture medium (in 2.5 ml of medium per 25-cm² flask) for 24 h before collecting cells and isolation of nuclei. Equal amounts of nuclei (5 μ g DNA) were incubated for 2 h in 20- μ l reactions containing *Xenopus* RNA polymerase III and labelled and unlabelled nucleoside triphosphates^{16,29}. RNA was isolated from these reactions and analysed on a denaturing polyacrylamide gel. Lanes: 1, 5S RNA marker; 12, tRNA^{tyr} gene marker; 2, cells not exposed to polyamide; 3–11, cells exposed to the indicated polyamides at the following concentrations: 100 nM in lanes 3, 6 and 9; 300 nM in lanes 4, 7, 10; 1 μ M in lanes 5, 8, 11. **b**, Graphic representation of the effects of polyamides **1**, **2** and **3** on 5S RNA and tRNA transcription expressed as the ratio of 5S RNA to tRNA gene transcription relative to the ratio obtained in the absence of polyamide. Error bars represent estimated standard deviation.

shows a graphical representation of the effect of the polyamides on 5S RNA gene transcription relative to transcription of the tRNA genes.

Concentrations of polyamide 1 as low as 100 nM have a pronounced and selective effect on 5S gene transcription. At higher polyamide concentration, there is a general decrease in the transcriptional activity of the nuclei; however, at each concentration tested, the effects of the polyamide are much greater on 5S RNA transcription than on tRNA transcription (Fig. 4a). Having established that nearly maximal inhibition of 5S transcription is achieved with 1 μ M polyamide 1, we monitored nuclear transcription after various times of cell growth in the presence of the polyamide. No inhibition is observed for zero-time incubation with polyamide 1 at 1 μ M, indicating that disruption of transcription complexes does not occur during or after the isolation or work-up of cell nuclei. Inhibition of 5S RNA gene transcription was statistically equivalent when cells were exposed to polyamide 1 for 24, 48 or 72 h.

These nuclear transcription experiments suggest that polyamide 1 is able to enter cells, transit to the nucleus and disrupt transcription complexes on the chromosomal 5S RNA genes. To rule out the possibility that the observed inhibition is due to some nonspecific toxicity of the polyamide rather than to direct binding to the 5S RNA gene, the effects of mismatch polyamides 2 and 3 in the nuclear transcription assay were monitored (Fig. 4a). Only a small effect on 5S RNA synthesis relative to tRNA synthesis is observed with 1 μ M of the mismatch polyamides 2 or 3 in the culture medium for 24 h. This result suggests that the general inhibition of transcription obtained at high concentrations of polyamide 1 (Fig. 4a) may be a secondary effect of the inhibition of 5S RNA synthesis *in vivo*, rather than the result of nonspecific polyamide interactions. Polyamide 2 effects a small enhancement of 5S RNA transcription *in vitro* and *in vivo*, indicating that polyamides may be able to upregulate transcription in certain cases.

Selective inhibition of RNA polymerase III transcription of a 5S RNA gene relative to a tRNA gene is a first step towards exploring the potential of polyamides as artificial regulators of eukaryotic gene expression. It remains to be determined whether a broad panel of genes can be selectively targeted within a variety of cell types. We have found that a polyamide binding in the minor groove adjacent to the TATA element is sufficient to inhibit transcription by RNA polymerase II (unpublished results). The optimal polyamide binding-site size required to elicit specific biological function, as well as strategies to inhibit major-groove-binding proteins with polyamides, will be described elsewhere. In addition to down-regulation of genes, polyamides might be designed specifically to upregulate transcription, thereby expanding the repertoire of synthetic cell-permeable transcription factors with potential clinical application. □

Methods

Polyamides. Polyamides were synthesized by solid-phase methods²⁵. The identity and purity of the polyamides was verified by ¹H NMR, matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS), and analytical HPLC. MALDI-TOF-MS: 1, 1223.4 (1223.3 calculated for M + H); 2, 1222.3 (1222.3 calculated for M + H); 3, 1223.1 (1223.3 calculated for M + H).

Transcription inhibition *in vitro*. A high-speed cytosolic extract from unfertilized *Xenopus* eggs was prepared as described²⁶. DNA templates for transcription were the somatic-type 5S RNA gene contained in plasmid pX1s11 (ref. 27; 50 ng per reaction) and the tyrD tRNA gene contained in plasmid pTyrD²⁴ (100 ng plasmid DNA per reaction), both from *X. laevis*. Transcription reactions (20 μ l final volume) contained: 2.5 μ l extract, 9 ng (12 nM) of TFIIIA isolated from immature oocytes²⁸, 0.6 mM ATP, UTP, CTP, 0.02 mM GTP and 10 μ Ci [α -³²P]GTP, and 12 mM HEPES (pH 7.5) buffer containing (final concentrations) 60 mM KCl, 6 mM MgCl₂, 25 μ M ZnCl₂ and 8% (v/v) glycerol. Plasmid DNAs were preincubated with polyamides in the same buffer before adding TFIIIA and other reaction components. RNA was purified and analysed on a denaturing 6% polyacrylamide gel. A Molecular Dynamics Phosphor-

imager equipped with ImageQuant software was used to quantify the effect of the polyamides on 5S and tRNA gene transcription.

Transcription inhibition *in vivo*. Fibroblasts from a *Xenopus*-kidney-derived cell line (kindly provided by P. Labhart) were grown at ambient temperature in 25-cm² culture flasks in Dulbecco's modified Eagle medium containing 10% (v/v) fetal calf serum. Cells were passaged for a minimum of three days before addition of polyamide to the culture medium. Incubations were continued for various times and nuclei were prepared by hypotonic lysis and used as templates for transcription as described¹⁶. DNA content was determined by measuring the absorbance of an aliquot of the isolated nuclei in 1% (w/v) sodium dodecyl sulphate (using an extinction coefficient at 260 nm of 1 absorbance unit for 50 μ g ml⁻¹ DNA). Buffer components and labelled and unlabelled nucleoside triphosphates were as for plasmid transcription. Reactions were supplemented with 2 μ l RNA polymerase III (at ~50 μ g ml⁻¹) isolated from *Xenopus* oocytes²⁹.

Received 24 January; accepted 3 March 1997.

1. Trauger, J. W., Baird, E. E. & Dervan, P. B. Recognition of DNA by designed ligands at subnanomolar concentrations. *Nature* **382**, 559–561 (1996).
2. Moser, H. E. & Dervan, P. B. Sequence-specific cleavage of double helical DNA by triple helix formation. *Science* **238**, 645–650 (1987).
3. Thuong, N. T. & Helene, C. Sequence-specific recognition and modification of double-helical DNA by oligonucleotides. *Angew. Chem. Int. Ed. Engl.* **32**, 666–690 (1993).
4. Maher, J. L., Dervan, P. B. & Wold, B. Analysis of promoter-specific repression by triple-helical DNA complexes in a eukaryotic cell-free transcription system. *Biochemistry* **31**, 70–81 (1992).
5. Duvalvalentin, G., Thuong, N. T. & Helene, C. Specific inhibition of transcription by triple helix-forming oligonucleotides. *Proc. Natl Acad. Sci. USA* **89**, 504–508 (1992).
6. Ho, S. N., Boyer, S. H., Schreiber, S. L., Danishefsky, S. J. & Crabtree, G. R. Specific inhibition of formation of transcription complexes by a calicheamicin oligosaccharide—a paradigm for the development of transcriptional antagonists. *Proc. Natl Acad. Sci. USA* **91**, 9203–9207 (1994).
7. Liu, C. *et al.* Sequence-selective carbohydrate DNA interaction—dimeric and monomeric forms of the calicheamicin oligosaccharide interfere with transcription factor function. *Proc. Natl Acad. Sci. USA* **93**, 940–944 (1996).
8. Wade, W. S., Mrksich, M. & Dervan, P. B. Design of peptides that bind in the minor groove of DNA at 5'-(A,T)G(A,T)C(A,T) sequences by a dimeric side-by-side motif. *J. Am. Chem. Soc.* **114**, 8784–8794 (1992).
9. Mrksich, M. *et al.* Antiparallel side-by-side dimeric motif for sequence-specific recognition in the minor groove of DNA by the designed peptide 1-methylimidazole-2-carboxamidenetropsin. *Proc. Natl Acad. Sci. USA* **89**, 7586–7590 (1992).
10. Pelton, J. G. & Wemmer, D. E. Structural characterization of a 2-1 distamycin A-d(CGCAAATTTG GC) complex by two-dimensional NMR. *Proc. Natl Acad. Sci. USA* **86**, 5723–5727 (1989).
11. White, S., Baird, E. E. & Dervan, P. B. Effects of the A/T/T/A degeneracy of pyrrole-imidazole polyamide recognition in the minor groove of DNA. *Biochemistry* **35**, 12532–12537 (1996).
12. Parks, M. E., Baird, E. E. & Dervan, P. B. Optimization of the hairpin polyamide design for recognition of the minor groove of DNA. *J. Am. Chem. Soc.* **118**, 6147–6152 (1996).
13. Trauger, J. W., Baird, E. E., Mrksich, M. & Dervan, P. B. Extension of sequence-specific recognition in the minor groove of DNA by pyrrole-imidazole polyamides to 9–13 base pairs. *J. Am. Chem. Soc.* **118**, 6160–6166 (1996).
14. Kelly, J. J., Baird, E. E. & Dervan, P. B. Binding site size limit of the 2:1 pyrrole-imidazole polyamide-DNA motif. *Proc. Natl Acad. Sci. USA* **93**, 6981–6985 (1996).
15. Geierstanger, B. H., Mrksich, M., Dervan, P. B. & Wemmer, D. E. Design of a G-C-specific minor groove-binding peptide. *Science* **266**, 646–650 (1994).
16. Schlissel, M. S. & Brown, D. D. The transcriptional regulation of *Xenopus* 5S RNA genes in chromatin—the roles of active stable transcription complexes and histone H1. *Cell* **37**, 903–913 (1984).
17. Engelke, D. R., Ng, S.-Y., Shastry, B. S. & Roeder, R. G. Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes. *Cell* **19**, 717–728 (1980).
18. Wolffe, A. P. & Brown, D. D. Developmental regulation of two 5S ribosomal RNA genes. *Science* **241**, 1626–1632 (1988).
19. Fairall, L., Rhodes, D. & Klug, A. Mapping of the sites of protection on a 5S-RNA gene by the *Xenopus* transcription factor-IIIa—a model for the interaction. *J. Mol. Biol.* **192**, 577–591 (1986).
20. Sakonju, S. & Brown, D. D. Contact points between a positive transcription factor and the *Xenopus* 5S RNA gene. *Cell* **31**, 395–405 (1982).
21. Clemens, K. R., Liao, X. B., Wolf, V., Wright, P. E. & Gottesfeld, J. M. Definition of the binding sites of individual zinc fingers in the transcription factor-IIIa-5S RNA gene complex. *Proc. Natl Acad. Sci. USA* **89**, 10822–10826 (1992).
22. Hayes, J. J. & Tullius, T. D. Structure of the TFIIIA 5S DNA complex. *J. Mol. Biol.* **227**, 407–417 (1992).
23. Clemens, K. R. *et al.* Relative contributions of the zinc fingers of transcription factor IIIA to the energetics of DNA binding. *J. Mol. Biol.* **244**, 23–35 (1994).
24. Stutz, E., Gouilloud, E. & Clarkson, S. G. Oocyte and somatic tyrosine transfer-RNA genes in *Xenopus laevis*. *Genes Dev.* **3**, 1190–1198 (1989).
25. Baird, E. E. & Dervan, P. B. Solid-phase synthesis of polyamides containing imidazole and pyrrole amino-acids. *J. Am. Chem. Soc.* **118**, 6141–6146 (1996).
26. Hartl, P., Gottesfeld, J. M. & Forbes, D. J. Mitotic repression of transcription *in vitro*. *J. Cell Biol.* **120**, 613–624 (1993).
27. Peterson, R. C., Doering, J. L. & Brown, D. D. Characterization of two *Xenopus* somatic DNAs and one minor oocyte specific 5S DNA. *Cell* **20**, 131–141 (1980).
28. Smith, D. R., Jackson, I. J. & Brown, D. D. Domains of the positive transcription factor specific for the *Xenopus* 5S-RNA gene. *Cell* **37**, 645–652 (1984).
29. Roeder, R. G. Multiple forms of DNA-dependent RNA-polymerases in *Xenopus laevis*—properties, purification, and subunit structure of class III RNA polymerases. *J. Biol. Chem.* **258**, 1932–1941 (1983).

Acknowledgements. We are grateful to the NIH for support of this work, the NSF for a predoctoral fellowship to J.W.T., and the Howard Hughes Medical Institute for a predoctoral fellowship to E.E.B.

Correspondence and requests for materials should be addressed to J.M.G. or P.B.D.

Crystal structure of the obese protein leptin-E100

Faming Zhang, Margret B. Basinski, John M. Beals, Stephen L. Briggs, Lisa M. Churgay, David K. Clawson, Richard D. DiMarchi, Thomas C. Furman, John E. Hale, Hansen M. Hsiung, Brigitte E. Schoner, Dennis P. Smith, Xing Y. Zhang, Jean-Pierre Wery & Richard W. Schervitz

Research Technologies and Proteins, Endocrine Research Divisions, Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, Indiana 46285, USA

Mutations in the *obese* gene (OB) or in the gene encoding the OB receptor (OB-R) result in obesity, infertility and diabetes in a variety of mouse phenotypes^{1–7}. The demonstration that OB protein (also known as leptin) can normalize body weight in *ob/ob* mice has generated enormous interest^{8–11}. Most human obesity does not appear to result from a mutant form of leptin: rather, serum leptin concentrations are increased and there is an apparent inability to transport it to the central nervous system (CNS)¹². Injection of leptin into the CNS of overfed rodents resistant to peripheral administration was found to induce biological activity¹³. Consequently, for the leptin to act as a weight-lowering hormone in human obesity, it appears that appropriate concentrations must be present in the CNS. This places a premium on understanding the structure of the hormone in order to design more potent and selective agonists. Here we report the crystal structure at 2.4 Å resolution of a human mutant OB protein (leptin-E100) that has comparable biological activity to wild type but which crystallizes more readily. The structure reveals a four-helix bundle similar to that of the long-chain helical cytokine family¹⁴.

Leptin has 67% sequence identity among such diverse species as human, gorilla, chimpanzee, orangutan, rhesus monkey, dog, cow, pig, rat and mouse (Fig. 1). However, there is no apparent sequence similarity with any other proteins, making modelling of the structure difficult. Wild-type human leptin (hOB) aggregates extensively so it cannot readily be crystallized. Fortunately, a single amino-acid substitution of Glu for Trp at position 100 results in a protein, leptin-E100, with comparable biological activity and dramatically improved solubility and propensity to crystallize. We have determined the crystal structure of this mutant OB protein using two heavy-atom derivatives combined with solvent flattening (Table 1).

The molecule is well ordered and most of the residues are visible in the electron density map (Fig. 2a, b), except for a loop region from Thr 27 to Gly 38, which appears to be disordered.

OB protein packs as a monomer in this hexagonal crystal form. The protein is an elongated molecule with approximate dimensions of 20 Å × 25 Å × 45 Å. It consists of four antiparallel α-helices (A, B, C and D), connected by two long crossover links (AB and CD) and one short loop (BC), arranged in a left-hand twisted helical bundle as predicted by threading analysis¹⁵ (Fig. 3a). The four α-helices comprise the residues: A, Pro 2–His 26; B, Leu 51–Ser 67; C, Arg 71–Lys 94; D, Ser 120–Ser 143. The last five residues in helix D (from Gln 139 to Ser 143) form a kinked 3₁₀ helix. A large hydrophobic cylindrical core parallel to the helical bundle axis is formed from the mostly conserved residues of the four α-helices that face each other. These conserved residues may be important for maintaining the structural integrity of the protein. The four most conserved regions observed between OB proteins, Asp 9 to Thr 16, Met 54 to Ala 59, Asp 79 to His 88 and Tyr 119 to Gln 130 are located within the four α-helices (Fig. 1). The four-helix bundle takes an unusual up–up–down–down fold which forms a two-layer packing of antiparallel helix pairs A and D against B and C. The skew angle Ω between the two layers (AD:BC) is about 20°. The interhelix crossing angles range between –152° and –161°. The residues from Leu 107 to Glu 115 form the distorted helix E, which bends sharply in the middle at Gly 111 and Gly 112. It is packed almost perpendicularly against the four helical bundle (87° to helix D). The two long connecting loops AB and CD wrap around the same side of the four-helix bundle BD face, with the AB crossover loop passing in front of helix D and overlying residues Gly 131 and Ser 132 on its surface. The interhelical angles and features of the long crossover loops in the leptin-E100 structure are similar to those found in the long-chain helical cytokine crystal structures, which include granulocyte colony-stimulating factor (G-CSF)¹⁶, leukaemia inhibitory factor (LIF)¹⁷, ciliary neurotrophic factor (CNTF)¹⁸ and human growth hormone (hGH)¹⁹.

Despite the absence of sequence similarity between leptin and other long-chain helical cytokines, there is a striking structural similarity. As shown in Fig. 3b, the four-helix bundles of leptin-E100, hGH, LIF and G-CSF are highly superimposable. Leptin-E100 superimposes on LIF (85 Cα atoms) with an r.m.s. deviation of 2.50 Å, on G-CSF (101 Cα atoms) (2.37 Å), and on hGH (85 Cα atoms) (1.95 Å). Genetic homology is also observed among helical cytokine genes that have common intron–exon boundaries¹⁴. These similarities extend to their respective receptors. The extracellular domain of the OB-R shows strong homology to the glycoprotein gp130 signal-transducing subunits²⁰ of the interleukin-6 (IL-6) receptor, the G-CSF receptor and the LIF receptor, which couple to the JAK/STAT signal-transduction pathway^{21,22}. The structural

Table 1 Diffraction data, phasing and refinement statistics

Diffraction data and MIR phasing		Total reflections	Unique data (completeness)	R_{sym}^* (%)	$R_{\text{cullis}}^\dagger$	Phasing power [‡]
Derivative	Resolution (Å)					
Native	2.4	45,300	8,408 (99.3)	5.3	–	–
K ₂ PtCl ₆	2.6	39,313	6,653 (99.7)	5.6	0.81	1.04
K ₂ OsCl ₆	2.6	26,186	6,562 (98.6)	6.5	0.86	0.99

Figure of merit of MIRAS from 20–3.0 Å: 0.36 (0.75 after solvent flattening)

Crystallographic refinement		Atoms protein/solvent	Average B protein/solvent	R-factor (%)§		Bonds (Å)	R.m.s.d. Angles (°)
Resolution (Å)	Reflections ($F > 1\sigma$)			R_{work}	R_{free}		
8–2.4	7,900	1,025/83	39.3/49.6	20.5	29.0	0.010	1.78

* $R_{\text{sym}} = \sum |I_i - \langle I \rangle| / \sum I_i$, where I_i is the intensity of an individual measurement, and $\langle I \rangle$ is the mean intensity of multiple observations of symmetry-related reflections.
 $^\dagger R_{\text{cullis}} = \sum ||F_{\text{ph}} - F_{\text{p}}| - F_{\text{ncal}}| / \sum |F_{\text{ph}} - F_{\text{p}}|$ defined for centric reflections, where F_{ph} , F_{p} and F_{ncal} are native, derivative and calculated heavy-atom structure factor amplitudes, respectively.
 ‡ Phasing power = r.m.s. (F_{h}/E), where F_{h} is the mean amplitude of the heavy-atom structure factor and E is the lack-of-closure error.
 § R-factor = $\sum ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum |F_{\text{obs}}|$, where $|F_{\text{obs}}|$ and $|F_{\text{calc}}|$ are the observed and calculated structure factor amplitudes respectively. R_{free} is calculated for a randomly chosen 10% of reflections omitted from refinement, and R_{work} is calculated for the remaining 90% of reflections included in refinement.

similarity of the ligands and the sequence similarity of their receptors suggest a similar ligand–receptor mode of binding.

Detailed comparison of the leptin structure reveals some basic points of variation from other cytokines. These differences represent the distinctive structural requirements for high-affinity interaction of leptin with its specific receptor. G-CSF, LIF and hGH all have pronounced kinks in the middle of helix A, D or B, respectively. The distinctive kinks serve to maximize the close contact area between the helices in these structures. OB protein has only a small kink at the last helical turn of helix D between Leu 139 and Glu 140. Another difference is that helix B in G-CSF, LIF and hGH is about two turns longer than in OB protein. In addition, G-CSF, LIF and hGH all have extra helices in the AB loop, whereas OB protein

does not. But OB protein does have a small distorted helical segment E in the CD loop, which is in approximately the same three-dimensional position as one of the small helices in the AB loop of G-CSF, LIF and hGH. Another cytokine, interferon- β , contains a 17-amino-acid α -helix which is in the CD loop, but this large helix is aligned with the bundle axis rather than being perpendicular to it as in OB protein, and represents an interferon-like subfamily²³. Thus OB protein has a distinctive CD loop whose E helix bears a similarity to the interferon-like cytokines in terms of location in the primary sequence and a similarity to the long-chain helical cytokine family in terms of three-dimensional position.

OB proteins contain two cysteine residues, Cys 96 and Cys 146, which form a disulphide bond between the C terminus of the

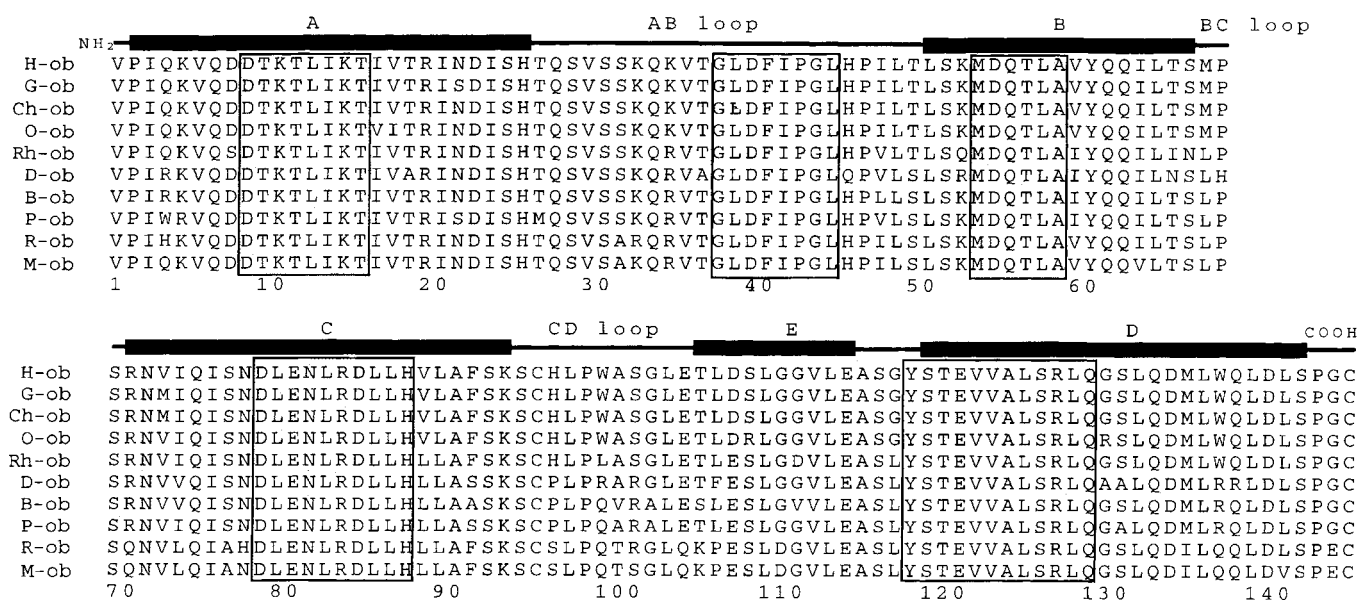


Figure 1 Sequence alignment of human, gorilla, chimpanzee, orangutan, rhesus monkey, dog, bovine, porcine, rat and mouse OB proteins. Secondary structure elements from the leptin-E100 crystal structure are indicated. Shadows highlight

the invariant residues in the OB protein family. The most conserved regions are boxed and shaded. There is a 67% sequence identity of OB protein between these species.

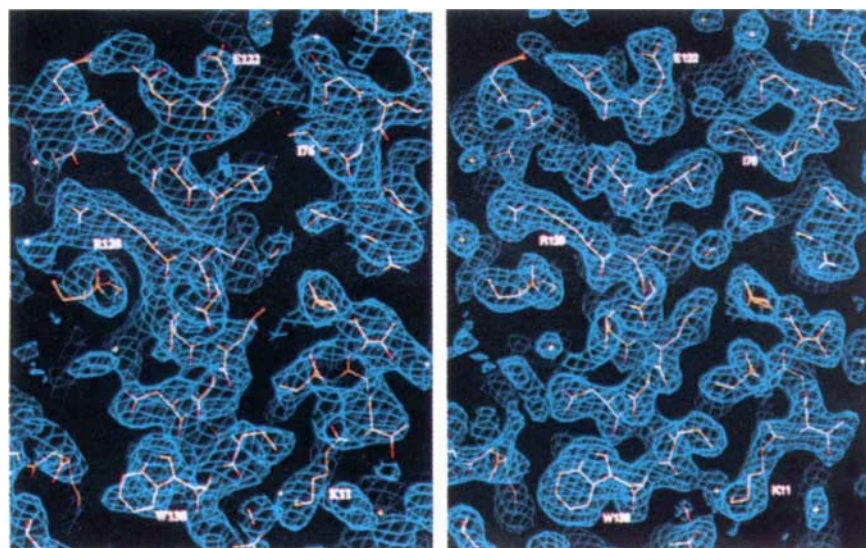


Figure 2 Electron density map contoured at 1 σ showing part of the antiparallel helices A and D with refined structure superimposed. **a**, Initial MIR map using two derivatives for data between 10 and 3.0 Å. **b**, (2 F_o - F_c) map calculated with

model-derived phases for data between 8.0 and 2.4 Å. Figures were drawn using program O (ref. 28).

protein and the beginning of the CD loop (Fig. 3a). This disulphide is exposed to the solvent on one end of the molecule, and it tethers the last turn of helix D nearly 36° towards the CD loop. The residues that comprise the kinked extension of helix D are Asn 139 to Ser 143 and are mostly conserved. Mutation of either of the two conserved cysteines renders the protein biologically inactive. This indicates that disulphide formation and the subsequent kinked D helix structure are important for structure folding and receptor binding.

The poor density observed for the first part of the AB loop, from residues Thr 27 to Gly 38, indicates high flexibility in this region. The hGH and hGH/hGHR complex structures show extensive conformational changes upon receptor binding in this region: from a coil in the free-ligand structure to a helix in the receptor complex¹⁹. Similarly, leptin, in the receptor-bound form, may also

assume an ordered conformation in this region. The rigidity of the four-helix-bundle core and the relative plasticity of the loop region has also been observed in other cytokine structures.

The CD loop in leptin-E100 is packed relatively tightly against the four-helical bundle and contains a small helical segment. Residues Leu 107 to Gly 111 form the first turn of helix E, which is followed by another distorted helical turn from Gly 112 to Gly 115. The angle between the two helical turns is $\sim 46^\circ$. Upon helix formation, Leu 104, Leu 107, Leu 110, Leu 114 and Val 113 in the CD loop pack tightly against the hydrophobic residues Val 60, Ile 64 and Met 68 in the C-terminal end of helix B and against Val 124 and Ala 125 in the N-terminal region of Helix D (Fig. 3c). Thus, helix E serves as a hydrophobic cap to bury the lipophilic residues on the surface of the BD helical bundle. Secondary structure analysis of IL-6 by

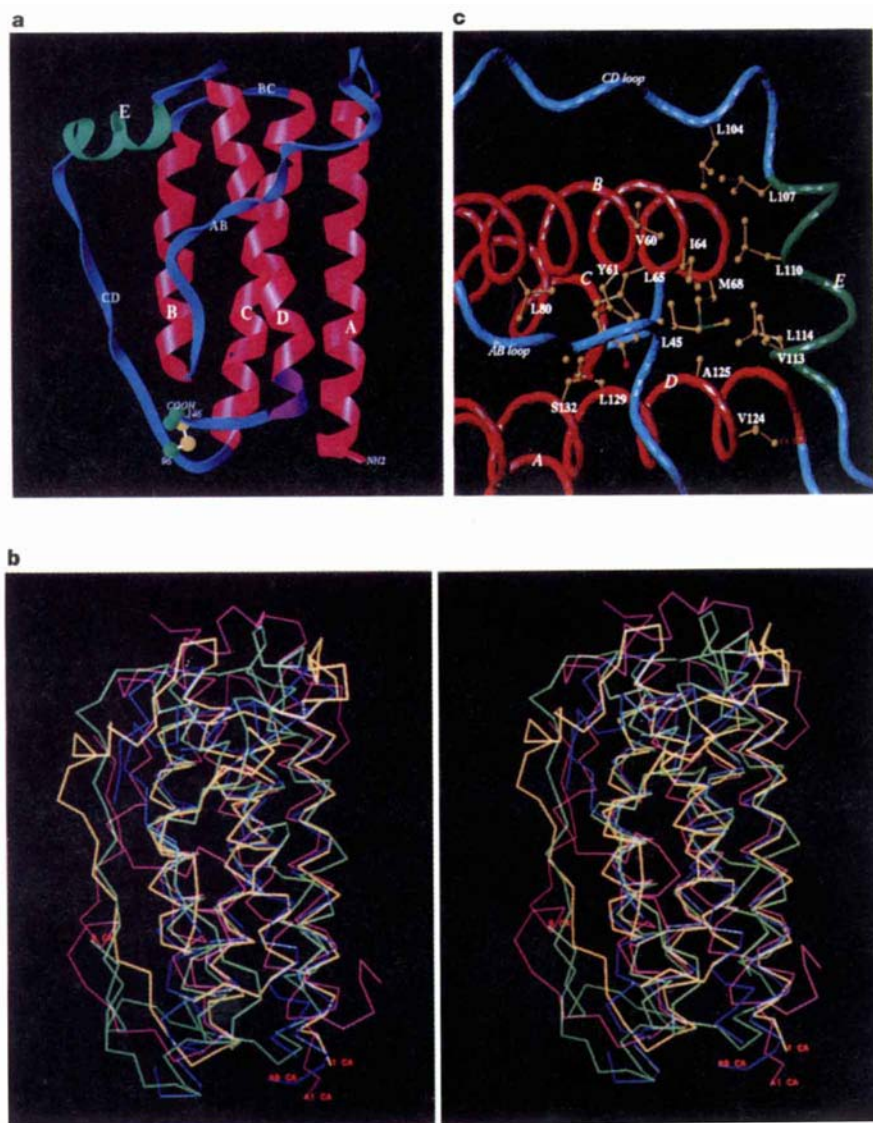


Figure 3 a, Ribbon diagram of OB protein. The view is perpendicular to the four-helical bundle axis. The four α -helices (A, B, C and D) are shown in red; the bent C-terminal 3_{10} helical extension of helix D is highlighted in magenta; the additional short helical segment E is green; the connecting loops AB, BC and CD are cyan. The N and C termini are indicated. The disulphide bridge between the C-terminal Cys 146 and Cys 96 in the AB loop is represented with ball-and-stick model. **b**, Stereo diagram of the C α superposition of OB protein (yellow) with other long-chain helical cytokines, hGH (magenta)¹⁹, LIF (green)¹⁷ and G-CSF (blue)¹⁶. Orientation as in **a**. The four-helical bundles superimpose well; the crossover connections are quite diverse in structure. **c**, Close-up of the hydrophobic patch

between helix E and helices B and D. The four-helical bundle is in red, the connecting loops are cyan, and the small distorted helix E is green. Conserved hydrophobic residues Leu 104, Leu 107, Leu 110, Leu 114 and Val 113 in the helical turn region pack against hydrophobic residues Val 60, Ile 64 and Met 68 in the C-terminal of helix B as well as Val 124 and Ala 125 in the N-terminal of helix D. Helix E serves as a hydrophobic cap for the B, D helices. Also shown is the buried Tyr 61 adjacent to the hydrophobic cap. The aromatic ring of Tyr 61 is packed against hydrophobic residues Leu 45, Leu 65, Leu 80, Ala 125 and Leu 129. Its hydroxyl group is hydrogen-bonded to the O $_B$ of Ser 132.

NMR also predicts an extra helix in the same region²⁴. Although the sequence of leptin in the CD loop region is less conserved among different species, the four leucines and Val 113 which are involved in the hydrophobic cap are highly conserved, with Leu 104, Leu 110, Leu 114 and Val 113 being invariant in ten species. Because of the distortion in this helix, not all of the usual hydrogen bonds are formed. The unsatisfied hydrogen bonds are compensated for by the hydrophobic interactions. The two glycines in the middle of the kink are partially conserved: at least one is present in all species of the OB protein studied. Thus, the helical bending at a central glycine may be an intrinsic structural feature of the OB protein motif.

It has been noticed that a buried Tyr or Phe aromatic side chain from helix D is well conserved among both short- and long-chain cytokines¹⁴. In leptin, there is only one buried aromatic residue, Tyr 61, and it is adjacent to the hydrophobic cap (Fig. 3c), but it is from helix B. The aromatic ring of Tyr 61 is buried in a hydrophobic pocket consisting of Leu 45, Leu 65, Leu 80, Ala 125 and Leu 129. The hydroxyl group of Tyr 61 is hydrogen-bonded to the O_β of Ser 132. Glu 100 lies on the surface of the molecule with its side chain pointing towards the solvent. Substitution of an exposed Trp by Glu at this position apparently reduces hydrophobic interactions and facilitates the crystallization of the OB protein.

The crystal structure of leptin-E100 suggests that it can be classified in the long-chain helical cytokine family, which also includes IL-6, IL-11, IL-12, LIF, G-CSF, CNTF and oncostatin M. The receptors of these cytokines belong to the IL-6 receptor subfamily which uses gp130 as a signal-transduction component in its receptor complex²⁵. Although induction by leptin of receptor oligomerization remains to be demonstrated, the structural similarity of these ligands may extend to their receptor-binding. The crystal structure of leptin-E100 provides a structural basis for understanding its biological activity and physical properties, which will serve as a foundation for the rational design of therapeutics with enhanced affinity and specificity while eliminating the inherent deficiencies of the native hormone. □

Methods

Diffraction data. Human OB protein containing the single amino-acid substitution of Glu for Trp 100 (leptin-E100) was crystallized by vapour diffusion (F.Z. *et al.*, manuscript in preparation). Crystals belong to space group P6₃ with cell dimensions $a = b = 88.14 \text{ \AA}$; $c = 48.05 \text{ \AA}$; $\alpha = \beta = 90^\circ$; $\gamma = 120^\circ$. For cryogenic data collection, crystals were transferred to stabilizing solution with 20% glycerol and soaked for 20 min, then flash-frozen in liquid N₂. X-ray data were collected at -170°C on a Raxis II imaging plate coupled to a Rigaku RU200 rotating anode X-ray generator. Data were processed using programs DENZO and SCALEPACK²⁶.

Multiple isomorphous replacement. Crystals were soaked in different solutions of heavy-atom compounds. Derivatized crystals were back-soaked in heavy-atom-free cryoprotecting solution for 20 min before data collection. The position of the platinum atom in the K₂PtCl₄ derivative (10 mM, 3 days) was determined from difference Patterson synthesis and MULTAN²⁷ difference maps. The relative position of the osmium atom in the K₂OsCl₆ derivative (5 mM, 2 days) with respect to the Pt site was located by difference Fourier synthesis using the Pt single isomorphous replacement phases plus anomalous dispersion. Heavy-atom parameters were refined by using MLPHARE in the CCP4 package²⁷.

Solvent flattening. The density of the crystal was measured using a carbon tetrachloride/*p*-xylene density-gradient column. The density is 2.43 g ml^{-1} , yielding a solvent content of 64% and consistent with one molecule per asymmetric unit. The initial electron-density map from two derivatives was calculated at $3.0 \text{ \AA}$. Phases were then modified using the iterative solvent-flattening, histogram-matching algorithm of DM²⁷. The solvent content was

conservatively set at 55%. The figure of merit was improved from 0.36 to 0.75 by this density modification procedure.

Model building and refinement. The polypeptide chain was fitted to the electron density with the program O²⁸ using the main-chain database and side-chain rotamer libraries. The initial model was then refined by the conventional positional refinement protocol in XPLOR²⁹, gradually extending the resolution to $2.8 \text{ \AA}$. Simulated annealing was then used at 2.8, 2.6 and $2.4 \text{ \AA}$ resolution, respectively. Cycles of partial phase combination in SIGMMA²⁹ were performed. Simulated annealing omit maps were used extensively for the model building of the weak loop region. The individual isotropic B-factors were refined in the last cycle. Residues from Thr 27 to Gly 38 have very poor electron density and high temperature factors, indicating high flexibility. They were omitted from the refinement. The final model contains 1,025 protein atoms and 83 ordered water molecules.

Received 14 February; accepted 24 March 1997.

1. Coleman, D. L. Obese and diabetes: two mutant genes causing diabetes-obesity syndromes in mice. *Diabetologia* **14**, 141–148 (1978).
2. Zhang, Y. *et al.* Positional cloning of the mouse *obese* gene and its human homologue. *Nature* **372**, 425–432 (1994).
3. Cioffi, J. A. *et al.* Novel B219/OB receptor isoforms: possible role of leptin in hematopoiesis and reproduction. *Nature Med.* **2**, 585–589 (1996).
4. Chehab, F. F., Mounzih, K., Lu, R. & Lim, M. E. Early onset of reproductive function in normal female mice treated with leptin. *Science* **275**, 88–90 (1996).
5. Ahima, R. S. *et al.* Role of leptin in the neuroendocrine response to fasting. *Nature* **382**, 250–252 (1996).
6. Tartaglia, L. A. *et al.* Identification and expression cloning of a leptin receptor, OB-R. *Cell* **83**, 1263–1271 (1995).
7. Lee, G.-H. *et al.* Abnormal splicing of the leptin receptor in diabetic mice. *Nature* **379**, 632–635 (1996).
8. Pellemounter, M. A. *et al.* Effects of the *obese* gene product on body weight regulation on *ob/ob* mice. *Science* **269**, 540–543 (1995).
9. Halaas, J. L. *et al.* Weight-reducing effects of the plasma protein encoded by the *obese* gene. *Science* **269**, 543–546 (1995).
10. Campfield, L. A. *et al.* Recombinant mouse OB protein: evidence for a peripheral signal linking adiposity and central neural networks. *Science* **269**, 546–549 (1995).
11. Stephens, T. W. *et al.* The role of neuropeptide Y in the antioesity action of the *obese* gene product. *Nature* **377**, 530–532 (1996).
12. Caro, J. F. *et al.* Decreased cerebrospinal-fluid/serum leptin ratio in obesity: a possible mechanism for leptin resistance. *Lancet* **348**, 159–161 (1996).
13. Heek, M. V. *et al.* Diet-induced obese mice develop peripheral, but not central, resistance to leptin. *J. Clin. Invest.* **99**, 385–390 (1997).
14. Sprang, S. R. & Bazan, J. F. Cytokine structural taxonomy and mechanisms of receptor engagement. *Curr. Opin. Struct. Biol.* **3**, 815–827 (1993).
15. Madej, T., Boguski, M. S. & Bryant, S. H. Threading analysis suggests that the *obese* gene product may be a helical cytokine. *FEBS Lett.* **373**, 13–18 (1995).
16. Hill, C. P., Osslund, T. D. & Eisenberg, D. The structure of granulocyte-colony-stimulating factor and its relationship to other growth factors. *Proc. Natl Acad. Sci. USA* **90**, 5176–5171 (1993).
17. Robinson, R. C. *et al.* The crystal structure and biological function of leukemia inhibitory factor: implication for receptor binding. *Cell* **77**, 1101–1116 (1994).
18. McDonald, N. Q., Panayotatos, N. & Hendrickson, W. A. Crystal structure of dimeric human ciliary neurotrophic factor determined by MAD phasing. *EMBO J.* **14**, 2689–2699 (1995).
19. De Vos, A. M., Ultsch, M. & Kossiakoff, A. A. Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science* **255**, 306–312 (1992).
20. Taniguchi, T. Cytokine signaling through nonreceptor protein tyrosine kinases. *Science* **268**, 251–255 (1995).
21. Vaisse, C. *et al.* Leptin activation of Stat3 in the hypothalamus of wild-type and *ob/ob* mice but not *db/db* mice. *Nature Genet.* **14**, 95–97 (1996).
22. Ghilardi, N. *et al.* Defective STAT signaling by the leptin receptor in diabetic mice. *Proc. Natl Acad. Sci. USA* **93**, 6231–6235 (1996).
23. Rozwarski, D. A. *et al.* Structural comparisons among the short-chain helical cytokines. *Structure* **2**, 159–173 (1994).
24. Xu, G.-Y. *et al.* Complete ¹H, ¹⁵N and ¹³C assignments, secondary structure, and topology of recombinant human interleukin-6. *J. Biomolecular NMR* **8**, 123–135 (1996).
25. Heldin, C.-H. Dimerization of cell surface receptors in signal transduction. *Cell* **80**, 213–223 (1995).
26. Otwinowski, Z. Oscillation data reduction program. *CCP4 Study Weekend* 29–30, 56–62 (1993).
27. Collaborative Computing Project, No. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
28. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
29. Brünger, A. T. *XPLOR Version 3.1: A System for X-ray Crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1993).

Acknowledgements. We thank M. Haslanger, J. Caro, W. Heath and T. Stephens for advice and encouragement; G. Becker, R. Chance, N. Chirgadze, F. Gu and A. Kline for discussion; and D. Allen, P. Atkinson, B. Coner, S. Dodd, L. Foster, C. Hanquier, J. Hoffmann, R. Millican, W. Muth, K. Ng, J. Rinella, E. Tinsley and M. Unger for their technical contribution.

Correspondence and requests for materials should be addressed to F.Z. (e-mail zhang_faming@lilly.com). Coordinates will be deposited in the Brookhaven Protein Data Bank.

Molecular biology marketplace

Browsing amongst the new molecular tools, the reader can find a high-throughput DNA sequencing and gene analysis system, electroelution capsules, a gel documentation workstation and a blunt PCR cloning system.

HP G2500A GeneArray scanner

From Hewlett-Packard

A genetic analysis system designed to scan Affymetrix GeneChip probe arrays

These single-use probe arrays have thousands of probe features that can sequence or identify multiple genes in a sample. In addition, different types of probe array applications are offered for specific, complex genetic analysis needs. The scanner itself is designed to eliminate the potential for human error associated with traditional genetic analysis methods. Featuring thermally compensated optics and low electronic noise, the scanner has the high sensitivity needed to read the small amounts of fluorescently labelled nucleic acid bound to probe arrays. According to the manufacturer, the system provides more genetic information per hour of technician time, higher sample throughput for improved sample processing and increased automation for less variability when contrasted with automated gel-based sequence analysis. Moreover, the scanner operates with few moving parts, has automated focusing and computer-controlled parameter settings for added reliability and ease of use. It precisely focuses a laser beam onto a 3- μm section of a 20- μm GeneChip. With a stated 5 min per scan, this unit can analyse probe arrays with thousands of base pairs of information, requiring little training and maintaining low operator dependence.

Reader Enquiry No. 100

OpenGene automated DNA sequencing

From Visible Genetics

An automated DNA sequencing system designed specifically for mutation detection in the clinical research laboratory

Based around an ultra-thin disposable cassette, the system provides speed and simplicity for automated sequencing and fragment analysis. According to the manufacturer, MicroCel cassettes are filled and



OpenGene cassette-based DNA sequencing.

polymerized in just over a stated three minutes, and 300 bases are sequenced in as little as 30 minutes. The system includes GeneObjects DNA analysis software and a complete line of gene-specific GeneKits, which are under development for applications that include infectious diseases, oncology and tissue typing.

Reader Enquiry No. 101

Cell Shock electroporator

From Hybaid

Developed for the cost-conscious laboratory wishing to transform both prokaryote and eukaryote cell types

The unit has been designed to be compact and self contained, with no external cuvette holding device, no connecting cables and no outside power supply. Safe operation is said to be assured through the use of a built-in safety interlock, which prevents pulse delivery until the cuvette cover is completely shut. The traditional requirement to set complex resistance and capacitance parameters has been replaced through the use of three simple pre-set programmes. Used in conjunction with either 1-, 2- or 4-mm electroporation cuvettes, no further adjustments are required to attain high transformation efficiencies, according to the manufacturer.

Reader Enquiry No. 102

IR² System

From Li-Cor

A high-throughput DNA sequencing system

This DNA sequencing and genetic analysis system features two infrared dyes that, say Li-Cor, greatly increase throughput, can save 50 per cent on enzyme usage and can cut labour in half. For DNA sequencing, the system is said to increase productivity by sequencing over 100 kb per day, with up to 64 samples per run. Faster data generation is said to be the result of consistent 800- to 1,200-base readlengths with 99 per cent accuracy, as well as simultaneous bi-directional sequencing strategies. For genetic analysis, the IR² System is designed to run up to 9,000 genotypes per day. Sizing and quantification is performed with Gene ImagIR software. Using Gene ImagIR, even dinucleotide repeats can be accurately resolved with true gel images from the IR² System. The software uses an integrated relational database to simplify project management, sample comparison and data output for linkage analysis. In addition to DNA sequencing and genotyp-



Li-Cor's IR² DNA sequencing and gene analysis system.

ing, other applications include mutation analysis, HLA typing, human identification, linkage, mapping and phylogenetic studies.

Reader Enquiry No. 103

DNA purification

DNA Etox reagent

From Sterogene Bioseparations

Formulated to remove endotoxin from plasmid and synthetic DNA intended for gene therapy applications

A high-efficiency, solid-phase reagent, DNA Etox has an extremely high affinity for endotoxin, with a binding capacity up to 1,000 EU ml⁻¹ as measured in a standard plasmid preparation. Affinity of the immobilized ligand for DNA is low, delivering a mass yield that typically exceeds 90 per cent. DNA Etox is readily regenerated with 1 M NaOH and can be sterilized by autoclaving. The reagent is used either in a batch or in a chromatography column.

Reader Enquiry No. 104

Micro BioSpin columns

From Bio-Rad

These ready-to-use columns work with small samples (20–75 μl) for quick removal of unincorporated dye terminators used in automated sequencing reactions

For use with all standard benchtop microcentrifuges, the Micro BioSpin columns have a rapid four-step protocol, which is stated to take just ten minutes. Retaining up



Micro BioSpin columns for samples up to 75 μl .

to 100 per cent of unincorporated nucleotides with a 95 per cent recovery of applied DNA, the columns are said to ensure efficient recovery and high purity for smaller sample volumes, involving no hydration, bubble removal or any further preparation. Gels are available in Tris buffer, which does not cause salt effects with fluorescent sequencing gels, or in SSC buffer. In addition to being fully autoclavable, the Micro BioSpin columns are specially sized with exclusion limits of either 5 or 20 base pairs. Applications for the columns include the removal of dye terminators, cleaning up DNA primers, removing radiolabels removing nucleotides, and desalting DNA and proteins.

Reader Enquiry No. 105

PhosphoMagnaCel

From Vector Laboratories

For rapid purification of DNA binding proteins or other positively charged proteins

Proteins can be separated from crude mixtures rapidly and simply with this product using a one-step, one-tube purification protocol. The derivatized magnetizable ion exchanger, developed by Scigen, consists of negatively charged phosphate groups attached to sugar residues on a paramagnetic support. The magnetic separation technique allows for easy, gentle manipulation of protein solutions, ensuring high yields of active protein. Unlike lengthy column chromatography procedures, minimum preparation time is required. This system is said to be well suited for the simultaneous processing of multiple samples.

Reader Enquiry No. 106

Biomek accessory

From Beckman

A new gripper tool for use with the 96 Plasmid Purification kit on the Biomek 2000 laboratory automation workstation

This gripper tool automatically moves plates on, off, into and out of the 96 Filtration System vacuum manifold, which is used along with the vacuum valve unit for DNA purification on the Biomek 2000 kit. The kit contains reagents, filter plates and



Beckman offers a new gripper tool for the BioMek2000 plasmid purification system.

collection plates for the purification of 96 templates. When used on the automated workstation with the new gripper tool, the kit's chemistry is completely automated, allowing the researcher to load bacterial pellets, walk away and return to purified templates in a 96-well plate. The kit produces DNA ready for use in fluorescent cycle sequencing reactions. It may also be used with manual procedures.

Reader Enquiry No. 107

QuikPik electroelution capsules

From Stratagene

Rapid recovery of DNA, RNA and protein from agarose or acrylamide gels

These electroelution capsules recover DNA, RNA and protein from gels with as little as 90 s hands-on time. This product is said to eliminate the need for glass milk, spin columns, binding columns or other methods that can damage DNA. This electroelution method is designed to eliminate washing, spinning, heating and precipitations that can also harm DNA. These capsules allow the user to recover rapidly 1,000 bp of DNA, up to 95 per cent. The method involves picking the band of interest with the punch, putting the punch in the cup and placing it on top of the submerged gel with the cup facing the flow of current. Then, the current is reapplied to move the sample out of the gel. The eluted sample is collected by punching a hole at the collection port and aspirating the sample. Recovered DNA is then ready for ligation, restriction enzyme digests, sequencing, amplification, random priming or other enzymatic reactions.

Reader Enquiry No. 108

PCR and probe cleanup

From Nucleon Biosciences

New DNA purification cartridges that have been designed to be simple to use and versatile

The cartridges are based on a 1.5-ml spin microfuge that contains a size-exclusion resin with a cut-off of 75 bp. Unlike competitive systems that bind the desired DNA product, which is later eluted off, these cartridges allow the direct passage of the product through the cartridge while retaining all the smaller unincorporated nucleotides and primers. The company states that this system requires only a 1- to 2-s pulse on a microcentrifuge for effective separation.

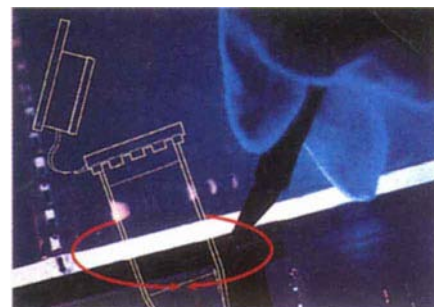
Reader Enquiry No. 109

Model RC10.10 vacuum concentrator

From Jouan

Standard nucleic acid and deluxe nucleic acid vacuum concentrator systems

Two new brochures describe the company's centrifugal vacuum concentrator systems



PCR and probe cleanup cartridges from Nucleon.

designed specifically for nucleic acid samples. One brochure describes the standard nucleic acid system and the other describes the deluxe nucleic acid system. Jouan's standard nucleic acid system is well suited to drying quickly DNA samples that have been precipitated using ethanol. The system comes complete with a high-capacity, Teflon-lined diaphragm pump, the model RC10.10 vacuum concentrator, an 84 × 1.5- to 2.0-ml rotor, an automatic vacuum bleeder valve, a chemical trap with molecular sieve cartridge and all necessary connecting tubing. There is no cold trap required with the system and the oil-free diaphragm pump is virtually maintenance free. The deluxe nucleic acid system includes the RC 10.22 with the patented pulse-vap system, which speeds drying times by a stated 25–40 per cent. According to the manufacturer, this system doubles sample throughput and allows faster, more efficient handling of DNA samples.

Reader Enquiry No. 110

RCM150 GX centrifuge

From Sorvall

An ultramicrocentrifuge that has been designed to provide an alternative to using columns for lipoprotein and plasmid DNA isolation in the laboratory

Capable of reaching 150,000 r.p.m. (almost 900,000g) in under 100 s, the 'whisper-quiet' micro-ultracentrifuge will separate lipoproteins and band plasmid DNA in as little as 2 h, delivering higher yields and purer samples in the process, according to the manufacturer. Samples loaded in the morning can be ready for use before lunch, and since overnight separations are eliminated, timed enzyme and protein experiments can be performed without loss of biological activity. The RCM-150 GX is small enough to fit into tight spots in the laboratory, saving valuable bench space. Measuring just 1,000 mm high × 490 mm deep × 440 mm wide, the instrument is stated to take up less than one quarter of the space occupied by conventional floor-standing ultracentrifuges. A quick-stepping rotor system eliminates the need for torque wrenches and buttons, while a highly advanced damper bearing allows tubes to be balanced visually, eliminating the need



Sorvall's RCM150GX — an alternative to columns for lipoprotein and plasmid isolation.

to weigh samples. Ten programmes, each with up to nine steps, can be stored into memory, and real-time control software permits easy scheduling for unattended runs. Run parameters and instrument status are displayed on a large LCD screen in a 'flip-back' panel, which closes to cover the soft-touch keypad when the instrument is not in use. The direct, brushless drive system eliminates the need for periodic maintenance and is warranted for five years with unlimited revolutions.

Reader Enquiry No. 111

Gel analysis hardware

PHD 2000

From Ultra-Lum

A gel documentation workstation designed for the multiuser laboratory

This system utilizes a 386 computer (486 upgrade optional) with special ten-key touchpad that provides simple, one-key operation for the enhancement, archiving, storage and printing of gels. The touchpad design prevents damage to keys from liquids or chemicals. A 3.5-inch floppy disk stores information in a selectable format and allows for analysis of gels on any remote Macintosh or IBM-compatible computer. The touchpad module can be installed in any existing IBM-type computer by using any open floppy drive slot. When combined with the Ultra-Lum FP 1000 low profile darkroom, a high-resolution CCD camera, a 23-cm monitor and a 265 grey-scale thermal printer, the PHD 2000 provides a cost-effective multi-user gel documentation workstation. It is useful for laboratories that run a variety of gel types, such as ethidium bromide, protein, coomassie blue, silver stains, northern and Southern blots, and other DNA gels.

Reader Enquiry No. 112

Chemiluminescent standards

From Autogen Bioclear

These are chemiluminescent molecular weight standards developed by Santa Cruz Biotechnology for western blotting

These standards are intended for use in conjunction with luminol-based detection reagents. The ladder consists of six bands (132 K, 90 K, 62 K, 43 K, 36 K and 23 K) that appear on western blot film when used in conjunction with Cruz Marker compatible western blotting secondary antibodies. Cruz Marker standards are provided in loading buffer and can be applied directly to SDS-PAGE gels. Anti-rabbit, -mouse, -rat and -goat secondary antibodies compatible with western blotting are also available, as is the luminol reagent.

Reader Enquiry No. 113

UV/white converter

From Ultra Violet Products

A phosphor conversion screen that will change the output from a normal ultraviolet transilluminator to white light

This converter should allow the user to view ethidium bromide gels using ultraviolet light, and protein gels using white light. Moreover, as the white light output is of a very high level, it is also well suited for viewing all types of transparent gel media. The converter is made of plastic, thus making it virtually unbreakable. It is mounted into a tray that can be lifted onto and off any transilluminator effortlessly. It can also cover 21 × 26-cm and 20 × 40-cm filter frames.

Reader Enquiry No. 114

Dark Reader

From Clare Chemical Research

A new apparatus for the detection of nucleic acids and proteins in gels

The dangers associated with the ultraviolet light emitted by transilluminators are known to cause eye and skin damage for the user, as well as chemical damage to DNA samples. Now, Clare Chemical Research has introduced a new transilluminator for viewing DNA and proteins in gels. It uses a low-powered fluorescent lamp to excite fluorophores, such as SYBR Green, ethidium bromide and SYPRO Orange in the visible spectrum. The Dark Reader transilluminator offers three modes: 'very safe' with essentially zero ultraviolet output; 'drastically reduced DNA damage' with the same sensitivity as ultraviolet for many dyes and allows viewing of DNA in gels while running; and a mode that is 100 per cent compatible with the photographic and video documentation system. The unit is currently available in two sizes: 21 cm × 14 cm and 42 × 28 cm.

Reader Enquiry No. 115

Assays and kits

CpG WIZ

From Oncor

A p16 methylation detection kit for the detailed examination of hypermethylated CpG islands and transcriptional inactivation

This PCR-based assay enables the specific detection of methylated CpG islands within the p16 promoter. The methodology is based on the chemical modification of genomic DNA with sodium bisulphite, and subsequent methylation specific PCR (MSP). Chemical modification with sodium bisulphite creates sequence differences between tumour and normal DNA by converting cytosine to uracil residues and leaving methylated cytosines unchanged. MSP is then performed using specific p16 gene primers, which are able to distinguish methylated from unmethylated samples. Applications for this assay include analysis of p16 gene inactivation in solid tumours and lymphomas, evaluation of clonal subpopulations in heterogeneous tumour samples, comparison of methylation changes and pathological analysis. It also allows enquiries into the stages of early tumour evolution and micrometastasis, the investigation of demethylating agents, which reverse hypermethylation and cell-cycle research.

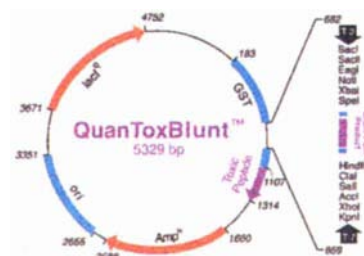
Reader Enquiry No. 116

Blunt PCR

From Quantum Biotechnologies

A positive selection system for easy Blunt PCR product cloning

The introduction of new thermostable polymerases that produce blunt-ended PCR products has created the need for new PCR products cloning methods. Current blunt-end PCR products cloning kits involve the use of rare restriction enzymes, special PCR strategies and/or expensive antibiotic selection. The GATA1 gene, present in the QuanToxBlunt vector, codes for a toxic peptide that inhibits the growth of bacterial cells when the expression is induced. Ligation of DNA fragments into the blunt cloning site disrupts the gene coding for the toxic peptide. The recombinant molecules are then non-toxic while vector molecules religated on themselves



Quantum Biotechnologies' QuanToxBlunt.

are toxic. This constitutes a positive selection vector ensuring a clean background, which is stated to produce more than 90 per cent positive recombinants. In order to produce and maintain the vector, the strong repressor *lac^d* gene is used and the presence of this gene in the vector allows the use of QuanToxBlunt in common bacterial strains. The vector is offered in linearized blunt-end form and extremities are controlled for absence of degradation.

Reader Enquiry No. 117

EasyFactor and PrimeFactor

From EquiBio

Two lipid transfection kits composed of polycationic lipids, together with a neutral, non-transfecting lipid compound

The manufacturer states that both kits have been used to transfect successfully a variety of cells including HeLa, Hep-2, CDS-7, Jurkat, keratinocytes and LNCaP cells. According to the manufacturer, EasyFactor has been found to be especially active for keratinocytes and tracheo-bronchial cells. PrimeFactor, in turn, has given high transfection efficiencies with liver cells. Their peak activities gave 13 pg and 26 pg of EasyFactor and PrimeFactor, respectively, for 1 g of DNA in 35-mm dishes. At this concentration, between 160 and 1,000 transfections can be performed with each millilitre of material. In addition to being more cost effective, the kits are also said to exhibit almost no significant toxicity when used at the optimal concentration.

Reader Enquiry No. 118

Hybrid Hunter System

From Invitrogen

A two-hybrid system specifically designed to simplify cloning and to function with any two-hybrid prey/library vector

Based on the interactive trap system, this system features the bait, pHybLex/Zeo, and the prey, pYESTrp. Both are said to be much smaller (4.8 and 5.8 kb, respectively), with improved multiple cloning sites for easier cloning and improved transformation efficiency. The bait plasmid, pHybLex/Zeo, also uses the novel antibiotic gene Zeocin for resistance. This allows the use of the system with other two-hybrid prey/library vectors.

Reader Enquiry No. 119

TNT T7 Quick System

From Promega

A coupled transcription and translation kit adds to the company's standard coupled TNT systems

All reagents supplied with this kit come optimized in a single tube to provide sim-



Invitrogen 1997 catalogue is also now available.

plified protocols. This is stated to reduce the time required to obtain *in vitro* translation products, when compared to standard rabbit reticulocyte lysate translations commonly using RNA synthesized *in vitro* from SP6, T3 or T7 RNA polymerase. The TNT T7 Quick systems take this one step further, combining the RNA polymerase, nucleotides, salts and RNase together with the reticulocyte lysate solution, forming a single master mix. This is said to offer maximum efficiency with minimum pipetting required, thus reducing the error rate.

Reader Enquiry No. 120

T4 Endonuclease V

From Cambio

T4 Endonuclease V was developed by Epicentre Technologies to detect damage in DNA caused by exposure to ultraviolet light

Transilluminators using ultraviolet light are commonly employed to visualize DNA in agarose or polyacrylamide gels. However, exposure to ultraviolet light for as little as 10 s is enough to damage DNA. The radiation produces covalent photoproducts, principally the *cis-syn* cyclobutane pyrimidine dimer (CPD). DNA that contains CPDs can be difficult, if not impossible, to clone and sequence. According to the manufacturer, the T4 Endonuclease V locates and binds to the CPD, cleaving the phosphodiester bonds next to it. This enables it to be used not only to detect, but also to quantify, the damage to the DNA. It should prove useful both for those who wish to test the integrity of samples in clone libraries, and for those wishing to determine why a particular DNA fragment will not clone.

Reader Enquiry No. 121

RNase Erase

From ICN Pharmaceuticals

RNase decontamination solution for complete removal of high RNase concentrations

This reagent is said to remove all traces of RNase contamination, eliminating the need to bake glassware or use DEPC solutions. Even trace quantities of RNase can lead to lower yields in *in vitro* transcription reactions, degradation during RNA purification protocols and variable results in northern blot transfers. This product contains three active ingredients to combat the presence of RNase and is said to be effective on glassware, plastic surfaces, countertops and pipettes. According to the manufacturer, it leaves no residues and will eliminate RNase contamination from microfuge tubes without inhibiting subsequent enzymatic reactions.

Reader Enquiry No. 122

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENTS

TIB
MOLBIOL

PNA

Order
per fax...
+49 30 78 79 94-99
or e-mail
DNA@TIB-MOLBIOL.DE

TIB is a service of Protein Engineering Institute

READER ENQUIRY NO. 86

NEW

LIPOSOMES

REPRODUCIBLE — HOMOGENEOUS —
EXTREMELY RAPID



Our new LIPOSOMAT guarantees rapid (5ml within 60 minutes) preparation of uniformly sized unilamellar liposomes. Liposome size is selectable between 25 and ca. 600 nm diameter. Sample volume is normally 5-10ml or up to 200ml if used in combination with an additional instrument. Lipid concentration can be up to 300 mg/ml.

LIPOSONORM

P.O. Box 650126, D-81215 Munich
For USA call 1-800 Liposome

READER ENQUIRY NO. 62